

STIC-Biotech/ChemLib

174171

From: Russel, Jeffrey  
Sent: Wednesday, December 14, 2005 9:09 AM  
To: STIC-Biotech/ChemLib  
Subject: Database Search Request, Serial Number: 10/789,494

RECEIVED  
DEC 14 2005  
STIC/BIOTECH/DIR/STIC

Requester:  
Jeffrey Russel (TC1600)  
Art Unit:  
1654  
Employee Number:  
62785  
Office Location:  
REM 3D19  
Phone Number:  
571-272-0969  
Mailbox Number:  
REM 3C18

Checked  
JRU  
12-20-2005

Case serial number:  
10/789,494  
Class / Subclass(es):  
NA  
Earliest Priority Filing Date:  
NA  
Format preferred for results:  
Diskette

Search Topic Information:

Please search SEQ ID NOS:3, 7, and 8 in the U.S. patent application sequence databases (pending, published, and issued) and in Geneseq/Uniprot/Pir. Thank you.  
Special Instructions and Other Comments:

\*\*\*\*\*

Searcher: \_\_\_\_\_  
Searcher Phone: \_\_\_\_\_  
Date Searcher Picked up: \_\_\_\_\_  
Date completed: \_\_\_\_\_  
Searcher Prep Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search  
NA# \_\_\_\_\_ AA# \_\_\_\_\_  
S/L: \_\_\_\_\_ Oligomer: \_\_\_\_\_  
Encode/Transl: \_\_\_\_\_  
Structure #: \_\_\_\_\_ Text: \_\_\_\_\_  
Inventor: \_\_\_\_\_ Litigation: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (Specify): \_\_\_\_\_

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2005, 23:08:17 ; Search time 143.75 Seconds  
(without alignments)  
70.301 Million cell updates/sec

Title: US-10-789-494B-3  
Perfect score: 109  
Sequence: 1 AASSVSSASSRSYDYSRNRNVRKN 23

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_21:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*  
9: geneseqp2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	109	100.0	23	8	ADU51229	Adu51229 Domestic
2	109	100.0	23	8	ADU51207	Adu51207 Silkworm
3	109	100.0	43	8	ADU51169	Adu51169 Domestic
4	73	67.0	654	9	ADZ09405	Adz09405 Canine pa
5	69	63.3	33	9	AEB30830	Aeb30830 Spider th
6	50	45.9	271	3	AAY57719	Aay57719 Spinach r
7	50	45.9	663	2	AAW29150	Aaw29150 Dual-spec
8	50	45.9	663	5	ABR52351	Abr52351 Protein r

9	50	45.9	663	5	ADI17164	Adi17164	Murine NO
10	50	45.9	1059	4	ABB71639	Abb71639	Drosophil
11	50	45.9	1059	8	ADS96722	Ads96722	Drosophil
12	49	45.0	383	2	AAR51059	Aar51059	Sequence
13	48	44.0	472	8	ADL27854	Adl27854	A vitis i
14	47.5	43.6	674	6	ABR53678	Abr53678	Protein s
15	47.5	43.6	674	7	ADK64078	Adk64078	Disease t
16	46	42.2	117	4	AAU32566	Aau32566	Novel hum
17	46	42.2	277	3	AAG20058	Aag20058	Arabidops
18	46	42.2	278	3	AAG20057	Aag20057	Arabidops
19	46	42.2	359	3	AAG20056	Aag20056	Arabidops
20	45	41.3	50	4	ABB22229	Abb22229	Protein #
21	45	41.3	50	4	AAM70055	Aam70055	Human bon
22	45	41.3	50	4	AAM57648	Aam57648	Human bra
23	45	41.3	50	4	ABG51748	Abg51748	Human liv
24	45	41.3	50	4	AAM05532	Aam05532	Peptide #
25	45	41.3	50	5	ABG39686	Abg39686	Human pep
26	45	41.3	146	7	ADG74666	Adg74666	Human kin
27	45	41.3	235	7	ADG74665	Adg74665	Human kin
28	45	41.3	317	4	ABG24218	Abg24218	Novel hum
29	45	41.3	388	7	ADC15643	Adc15643	E. intest
30	45	41.3	455	8	ABM83064	Abm83064	Human dia
31	45	41.3	455	8	ABM83061	Abm83061	Human dia
32	45	41.3	486	7	ADE58782	Ade58782	Human Pro
33	45	41.3	486	8	ADI82554	Adi82554	Human mod
34	45	41.3	486	8	ADR97409	Adr97409	Human PPP
35	45	41.3	486	9	ADX08183	Adx08183	Cyclin-de
36	45	41.3	487	8	ABM83063	Abm83063	Human dia
37	45	41.3	487	8	ABM83062	Abm83062	Human dia
38	45	41.3	892	7	ADD18910	Add18910	Human dis
39	45	41.3	916	7	ADB64143	Adb64143	Human pro
40	45	41.3	1002	7	ADC15645	Adc15645	E. intest
41	45	41.3	1298	7	ADF60378	Adf60378	Human con
42	45	41.3	1807	8	ADO17160	Ado17160	Novel exp
43	44.5	40.8	1677	8	ADO17159	Ado17159	Novel exp
44	44	40.4	91	5	ADK34170	Adk34170	Novel hum
45	44	40.4	215	8	ADH75876	Adh75876	Lucerne C

#### ALIGNMENTS

##### RESULT 1

ADU51229

ID ADU51229 standard; peptide; 23 AA.

XX

AC ADU51229;

XX

DT 24-FEB-2005 (first entry)

XX

DE Domestic silkworm fibroin H-chain peptide fragment 14.

XX

KW vulnerary; cell proliferation; wound healing; cell adhesion; cosmetics;  
KW cell culture; fibroin.

XX

OS Bombyx mori.

XX

PN JP2004339189-A.  
 XX  
 PD 02-DEC-2004.  
 XX  
 PF 04-DEC-2003; 2003JP-00406608.  
 XX  
 PR 28-FEB-2003; 2003JP-00055048.  
 XX  
 PA (DOKU-) DOKURITSU GYOSEI HOJIN NOGYO SEIBUTSU SH.  
 PA (TSUB/) TSUBOUCHI K.  
 XX  
 DR WPI; 2004-827614/82.  
 XX  
 PT New peptide having excellent cell growth promoting activity, for use as a  
 PT cell growth promoter, cell adhesion agent, wound healing-promoting agent,  
 PT cosmetic and cell culture base material.  
 XX  
 PS Example 3; Page; 27pp; Japanese.  
 XX  
 CC The invention relates to a novel peptide having excellent cell growth  
 CC promoting activity. The peptide of the invention demonstrates vulnerary  
 CC activity and may be utilised as a cell growth promoter, cell adhesion  
 CC agent, wound healing-promoting agent or cosmetic and cell culture base  
 CC material. The current sequence is that of a domestic silkworm fibroin H-  
 CC chain peptide fragment of the invention which is described as being  
 CC amorphous.  
 XX  
 SQ Sequence 23 AA;

Query Match 100.0%; Score 109; DB 8; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-10;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AASSVSSASSRSYDYSRRNVRKN 23  
 |||||  
 Db 1 AASSVSSASSRSYDYSRRNVRKN 23

#### RESULT 4

ADZ09405

ID ADZ09405 standard; protein; 654 AA.

XX

AC ADZ09405;

XX

DT 30-JUN-2005 (first entry)

XX

DE Canine parvovirus VP2 protein Seq 8.

XX

KW vaccine; antigen; VP2; silk; immune stimulation; antibody production;

KW virucide; antiparasitic; cytostatic; immunostimulant.

XX

OS Canine parvovirus.

XX

PN JP2005097229-A.

XX

PD 14-APR-2005.

XX  
 PF 03-DEC-2003; 2003JP-00405124.  
 XX  
 PR 04-DEC-2002; 2002JP-00352084.  
 PR 05-SEP-2003; 2003JP-00314808.  
 XX  
 PA (TORA ) TORAY IND INC.  
 XX  
 DR WPI; 2005-299766/31.  
 DR N-PSDB; ADZ09400.  
 XX  
 PT Vaccine for preventing infection caused by virus, parasite, and tumor,  
 PT comprises antigenic protein and component derived from silk thread.  
 XX  
 PS Disclosure; SEQ ID NO 12; 21pp; Japanese.  
 XX  
 CC This invention relates to a novel vaccine. Specifically, it refers to a  
 CC vaccine that comprises an antigenic protein, in particular the canine-  
 CC parvovirus VP2 protein and a component derived from silk thread. The  
 CC present invention describes a recombinant silkworm engineered to produce  
 CC this antigenic protein in a silk gland or silk thread, whereby the gene  
 CC encoding the antigenic protein has been transduced into a chromosome  
 CC using a vector to transfer this foreign gene. Administration of the  
 CC vaccine activates the immune response and so induces production of  
 CC antibodies against the antigenic canine-parvovirus VP2 protein in the  
 CC blood of an animal. Note that the silk thread containing this antigenic  
 CC protein is pulverized to form particles or is solubilized, hence the  
 CC vaccine can be obtained with high purity in a cost effective manner.  
 CC Accordingly, compositions derived thereof exhibit virucide,  
 CC antiparasitic, cytostatic and immunostimulant activities. This  
 CC polypeptide sequence is the canine parvovirus VP2 protein of the  
 CC invention. NOTE: This protein is given as an embedded protein the  
 CC sequence listing and is not further referred to in the specification.  
 XX  
 SQ Sequence 654 AA;

Query Match 67.0%; Score 73; DB 9; Length 654;  
 Best Local Similarity 100.0%; Pred. No. 0.0097;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 SRSYDYSRRNVRKN 23  
 |||||  
 Db 621 SRSYDYSRRNVRKN 634

RESULT 5  
 AEB30830  
 ID AEB30830 standard; peptide; 33 AA.  
 XX  
 AC AEB30830;  
 XX  
 DT 06-OCT-2005 (first entry)  
 XX  
 DE Spider thread peptide #1.  
 XX  
 KW Silk; spider thread protein.  
 XX

OS Bombyx mori.  
 XX  
 PN WO2005068495-A1.  
 XX  
 PD 28-JUL-2005.  
 XX  
 PF 12-JAN-2005; 2005WO-JP000619.  
 XX  
 PR 13-JAN-2004; 2004JP-00005489.  
 XX  
 PA (TORA ) TORAY IND INC.  
 PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
 XX  
 PI Hiramatsu S, Moriyama H, Asaoka R, Morita K, Tanaka T, Yamada K;  
 PI Obrien JP, Fahnestock SR;  
 XX  
 DR WPI; 2005-522809/53.  
 XX  
 PT Silk thread useful for producing textile fabric and in aeronautical  
 PT navigation, space exploration, has spider thread protein, produced by  
 PT transducing gene encoding spider thread protein to silkworm having  
 PT fibroin H-chain gene.  
 XX  
 PS Claim 16; SEQ ID NO 3; 48pp; Japanese.  
 XX  
 CC The invention relates to a silk thread comprising a spider thread  
 CC protein, produced by a transducing gene encoding spider thread protein in  
 CC a silkworm having a fibroin H-chain gene, without damaging the silkworm  
 CC fibroin H-chain gene. The invention also relates to producing silk thread  
 CC involving producing a transgenic silkworm and extracting silk thread from  
 CC the transgenic silkworm. The silk thread is useful for producing a  
 CC textile fabric and also useful in aeronautical navigation, space  
 CC exploration, to produce clothing, towrope and medical thread, etc. The  
 CC silk thread has high strength and elongation property. This sequence  
 CC represents a spider thread peptide of the invention.  
 XX  
 SQ Sequence 33 AA;

Query Match 63.3%; Score 69; DB 9; Length 33;  
 Best Local Similarity 100.0%; Pred. No. 0.0013;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 RSYDYSRRNVRKN 23  
 |||||  
 Db 1 RSYDYSRRNVRKN 13

Search completed: December 16, 2005, 23:46:08  
 Job time : 146.75 secs

GenCore version 5.1.6  
 Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2005, 23:40:43 ; Search time 34.9107 Seconds  
 (without alignments)  
 54.469 Million cell updates/sec

Title: US-10-789-494B-3  
 Perfect score: 109  
 Sequence: 1 AASSVSSASSRSYDYSRRNVRKN 23

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : Issued Patents\_AA:\*  
 1: /cgn2\_6/ptodata/1/iaa/5\_COMB.pep:\*  
 2: /cgn2\_6/ptodata/1/iaa/6\_COMB.pep:\*  
 3: /cgn2\_6/ptodata/1/iaa/H\_COMB.pep:\*  
 4: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
 5: /cgn2\_6/ptodata/1/iaa/RE\_COMB.pep:\*  
 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	50	45.9	434	2	US-09-270-767-61286	Sequence 61286, A
2	50	45.9	1059	2	US-09-270-767-45764	Sequence 45764, A
3	46.5	42.7	93	2	US-09-270-767-60930	Sequence 60930, A
4	46.5	42.7	427	2	US-09-270-767-45426	Sequence 45426, A
5	45	41.3	908	2	US-09-949-016-7580	Sequence 7580, Ap
6	45	41.3	916	2	US-10-104-047-2297	Sequence 2297, Ap
7	44	40.4	490	2	US-09-976-594-314	Sequence 314, Appl
8	44	40.4	490	2	US-09-905-999-23	Sequence 23, Appl
9	44	40.4	508	2	US-09-949-016-8519	Sequence 8519, Ap
10	44	40.4	508	2	US-09-949-016-8520	Sequence 8520, Ap
11	44	40.4	638	2	US-10-104-047-2626	Sequence 2626, Ap
12	43.5	39.9	235	2	US-09-248-796A-24586	Sequence 24586, A
13	43	39.4	172	2	US-09-252-991A-20172	Sequence 20172, A
14	43	39.4	223	2	US-09-270-767-38042	Sequence 38042, A
15	43	39.4	223	2	US-09-270-767-53259	Sequence 53259, A
16	42	38.5	141	2	US-09-252-991A-26722	Sequence 26722, A
17	42	38.5	168	2	US-09-107-532A-3748	Sequence 3748, Ap
18	42	38.5	272	2	US-09-107-532A-6938	Sequence 6938, Ap

19	42	38.5	353	1	US-08-687-702-37	Sequence 37, Appl
20	42	38.5	412	2	US-09-027-064-4	Sequence 4, Appli
21	42	38.5	412	2	US-09-271-815-4	Sequence 4, Appli
22	42	38.5	605	2	US-09-538-092-375	Sequence 375, App
23	42	38.5	2090	2	US-09-538-092-1081	Sequence 1081, Ap
24	42	38.5	2120	2	US-09-949-016-9768	Sequence 9768, Ap
25	41.5	38.1	185	2	US-09-248-796A-27129	Sequence 27129, A
26	41.5	38.1	246	2	US-09-104-678A-2	Sequence 2, Appli
27	41.5	38.1	247	2	US-09-548-938A-12	Sequence 12, Appl
28	41	37.6	224	1	US-08-248-466B-8	Sequence 8, Appli
29	41	37.6	259	2	US-09-270-767-45089	Sequence 45089, A
30	41	37.6	285	2	US-08-482-085B-20	Sequence 20, Appl
31	41	37.6	365	2	US-09-270-767-44316	Sequence 44316, A
32	41	37.6	380	2	US-09-270-767-43626	Sequence 43626, A
33	41	37.6	394	2	US-09-769-787-41	Sequence 41, Appl
34	41	37.6	528	2	US-09-270-767-61237	Sequence 61237, A
35	41	37.6	530	2	US-09-949-016-9519	Sequence 9519, Ap
36	41	37.6	648	1	US-08-817-436A-2	Sequence 2, Appli
37	41	37.6	670	2	US-09-270-767-45719	Sequence 45719, A
38	41	37.6	1032	2	US-09-489-039A-14025	Sequence 14025, A
39	41	37.6	1046	2	US-09-538-092-445	Sequence 445, App
40	40.5	37.2	100	2	US-09-513-999C-6133	Sequence 6133, Ap
41	40.5	37.2	238	2	US-09-632-570-15	Sequence 15, Appl
42	40.5	37.2	238	2	US-09-632-575-45	Sequence 45, Appl
43	40.5	37.2	239	2	US-09-216-295-15	Sequence 15, Appl
44	40	36.7	119	2	US-08-890-865A-17	Sequence 17, Appl
45	40	36.7	148	2	US-09-252-991A-28329	Sequence 28329, A

# ALIGNMENTS

## RESULT 1

US-09-270-767-61286  
; Sequence 61286, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 61286  
; LENGTH: 434  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-270-767-61286

Query Match 45.9%; Score 50; DB 2; Length 434;  
Best Local Similarity 45.5%; Pred. No. 8.7;  
Matches 10; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Qy 1 AASSVSSASSRSYDYSRRNVRK 22  
:|:| || :: | |||:|:  
Db 360 SANSKSSGGNKMSDVSRRNIRE 381



Search completed: December 17, 2005, 00:12:07  
Job time : 35.9107 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2005, 23:52:28 ; Search time 121.571 Seconds  
(without alignments)  
79.049 Million cell updates/sec

Title: US-10-789-494B-3  
Perfect score: 109  
Sequence: 1 AASSVSSASSRSYDYSRRNVRKN 23

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA\_Main:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	109	100.0	23	5	US-10-789-494B-3 Sequence 3, Appli
2	109	100.0	23	5	US-10-789-494B-59 Sequence 59, Appl
3	109	100.0	43	5	US-10-789-494B-21 Sequence 21, Appl
4	50	45.9	663	4	US-10-072-012-700 Sequence 700, App
5	50	45.9	663	5	US-10-029-345A-40 Sequence 40, Appl

6	50	45.9	1059	6	US-11-097-143-41709	Sequence 41709, A
7	48	44.0	1030	4	US-10-437-963-158754	Sequence 158754,
8	48	44.0	1507	4	US-10-425-115-342056	Sequence 342056,
9	47	43.1	181	4	US-10-767-701-45386	Sequence 45386, A
10	47	43.1	374	4	US-10-437-963-115405	Sequence 115405,
11	46	42.2	459	4	US-10-437-963-169698	Sequence 169698,
12	45	41.3	50	3	US-09-864-761-37527	Sequence 37527, A
13	45	41.3	208	4	US-10-437-963-184293	Sequence 184293,
14	45	41.3	211	4	US-10-437-963-159894	Sequence 159894,
15	45	41.3	317	5	US-10-450-763-54577	Sequence 54577, A
16	45	41.3	344	4	US-10-131-410-177	Sequence 177, App
17	45	41.3	458	4	US-10-437-963-184282	Sequence 184282,
18	45	41.3	916	4	US-10-104-047-2297	Sequence 2297, Ap
19	45	41.3	947	5	US-10-732-923-4287	Sequence 4287, Ap
20	44.5	40.8	177	4	US-10-425-115-362590	Sequence 362590,
21	44	40.4	123	4	US-10-767-701-43958	Sequence 43958, A
22	44	40.4	193	5	US-10-732-923-3690	Sequence 3690, Ap
23	44	40.4	215	4	US-10-421-138A-316	Sequence 316, App
24	44	40.4	286	3	US-09-864-761-46689	Sequence 46689, A
25	44	40.4	326	4	US-10-767-701-43959	Sequence 43959, A
26	44	40.4	357	4	US-10-437-963-151972	Sequence 151972,
27	44	40.4	436	4	US-10-282-122A-65292	Sequence 65292, A
28	44	40.4	490	3	US-09-905-999-23	Sequence 23, Appl
29	44	40.4	490	4	US-10-267-502-354	Sequence 354, App
30	44	40.4	490	5	US-10-825-177-23	Sequence 23, Appl
31	44	40.4	490	6	US-11-029-470-10	Sequence 10, Appl
32	44	40.4	534	6	US-11-097-143-17595	Sequence 17595, A
33	44	40.4	638	4	US-10-104-047-2626	Sequence 2626, Ap
34	44	40.4	1987	4	US-10-437-963-160584	Sequence 160584,
35	43.5	39.9	92	4	US-10-425-115-349762	Sequence 349762,
36	43.5	39.9	1143	4	US-10-032-585-7789	Sequence 7789, Ap
37	43	39.4	129	4	US-10-425-115-242444	Sequence 242444,
38	43	39.4	166	4	US-10-424-599-214039	Sequence 214039,
39	43	39.4	209	4	US-10-767-701-39778	Sequence 39778, A
40	43	39.4	222	4	US-10-424-599-254434	Sequence 254434,
41	43	39.4	222	5	US-10-739-930-8688	Sequence 8688, Ap
42	43	39.4	222	5	US-10-732-923-16338	Sequence 16338, A
43	43	39.4	223	4	US-10-425-114-42487	Sequence 42487, A
44	43	39.4	223	5	US-10-732-923-16168	Sequence 16168, A
45	43	39.4	223	5	US-10-732-923-16337	Sequence 16337, A

#### ALIGNMENTS

##### RESULT 1

US-10-789-494B-3

; Sequence 3, Application US/10789494B

; Publication No. US20050143296A1

; GENERAL INFORMATION:

; APPLICANT: TSUBOUCHI, Kozo

; APPLICANT: YAMADA, Hiromi

; TITLE OF INVENTION: EXTRACTION AND UTILIZATION OF CELL

; TITLE OF INVENTION: GROWTH-PROMOTING PEPTIDES FROM SILK PROTEIN

; FILE REFERENCE: OPS 635

; CURRENT APPLICATION NUMBER: US/10/789,494B

; CURRENT FILING DATE: 2004-02-27

; PRIOR APPLICATION NUMBER: JP 2003-55048  
; PRIOR FILING DATE: 2003-02-28  
; NUMBER OF SEQ ID NOS: 85  
; SEQ ID NO 3  
; LENGTH: 23  
; TYPE: PRT  
; ORGANISM: Bombyx mori  
US-10-789-494B-3

Query Match 100.0%; Score 109; DB 5; Length 23;  
Best Local Similarity 100.0%; Pred. No. 3.3e-09;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AASSVSSASSRSYDYSRRNVRKN 23  
| | | | | | | | | | | | | | | | | |  
Db 1 AASSVSSASSRSYDYSRRNVRKN 23

#### RESULT 4

US-10-072-012-700

; Sequence 700, Application US/10072012  
; Publication No. US20040033493A1  
; GENERAL INFORMATION:  
; APPLICANT: Tchernev, Velizar  
; APPLICANT: Spytek, Kimberly  
; APPLICANT: Zerhusen, Bryan  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Li, Li  
; APPLICANT: Gangolli, Esha  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Anderson, David W.  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Miller, Charles E.  
; APPLICANT: Gerlach, Valerie  
; APPLICANT: Taupier Jr, Raymond J.  
; APPLICANT: Gusev, Vladimir Y.  
; APPLICANT: Colman, Steven D.  
; APPLICANT: Wolenc, Adam R.  
; APPLICANT: Pena, Carol E. A  
; APPLICANT: Furtak, Katarzyna  
; APPLICANT: Grosse, William M.  
; APPLICANT: Alsobrook II, John P.  
; APPLICANT: Lepley, Denise M.  
; APPLICANT: Rieger, Daniel K.  
; APPLICANT: Burgess, Catherine E.  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-258  
; CURRENT APPLICATION NUMBER: US/10/072,012  
; CURRENT FILING DATE: 2002-01-31  
; PRIOR APPLICATION NUMBER: 60/265,102  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: 60/265,514  
; PRIOR FILING DATE: 2001-01-31  
; PRIOR APPLICATION NUMBER: 60/265,517  
; PRIOR FILING DATE: 2001-01-31

```

; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 700
;   LENGTH: 663
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-10-072-012-700

```

```

Query Match          45.9%; Score 50; DB 4; Length 663;
Best Local Similarity 57.1%; Pred. No. 88;
Matches 12; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

```

```

Qy      1 AASSVSSASSRSYDYSRRNVR 21
        ::|| ||:|| | | ||:||
Db      588 SSSSSSSSSSSSDLRRRDVR 608

```

```

RESULT 11
US-10-437-963-169698
; Sequence 169698, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules
Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 169698
;   LENGTH: 459
;   TYPE: PRT

```

; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_68093C.1.pep  
US-10-437-963-169698

Query Match 42.2%; Score 46; DB 4; Length 459;  
Best Local Similarity 39.1%; Pred. No. 2.3e+02;  
Matches 9; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

Qy 1 AASSVSSASSRSYDYSRRNVRKN 23  
|||:|::| |::|  
Db 36 AASSLSASASLDADFDKKQFRHN 58

Search completed: December 17, 2005, 00:17:32  
Job time : 122.571 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2005, 23:53:24 ; Search time 6.16071 Seconds  
(without alignments)  
25.151 Million cell updates/sec

Title: US-10-789-494B-3  
Perfect score: 109  
Sequence: 1 AASSVSSASSRSYDYSRRNVRKN 23

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 51470 seqs, 6736768 residues

Total number of hits satisfying chosen parameters: 51470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA\_New:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	44	40.4	436	6	US-10-467-657-4550	Sequence 4550, Ap
2	43	39.4	222	6	US-10-667-295-143	Sequence 143, App
3	39	35.8	131	6	US-10-667-295-37	Sequence 37, Appl
4	39	35.8	230	6	US-10-667-295-86	Sequence 86, Appl
5	39	35.8	277	6	US-10-667-295-85	Sequence 85, Appl
6	39	35.8	287	7	US-11-174-150-30	Sequence 30, Appl
7	39	35.8	483	6	US-10-632-150-48	Sequence 48, Appl
8	39	35.8	483	7	US-11-073-457-48	Sequence 48, Appl
9	39	35.8	483	7	US-11-073-460-48	Sequence 48, Appl
10	38.5	35.3	354	6	US-10-467-657-4352	Sequence 4352, Ap
11	38	34.9	152	6	US-10-957-569-38	Sequence 38, Appl
12	38	34.9	414	6	US-10-131-826A-418	Sequence 418, App
13	38	34.9	422	7	US-11-022-562-234	Sequence 234, App
14	37.5	34.4	477	7	US-11-089-551A-34	Sequence 34, Appl
15	37	33.9	227	7	US-11-170-653-27	Sequence 27, Appl
16	37	33.9	454	7	US-11-089-551A-35	Sequence 35, Appl
17	37	33.9	668	6	US-10-995-561-619	Sequence 619, App
18	37	33.9	691	6	US-10-995-561-617	Sequence 617, App
19	37	33.9	742	6	US-10-995-561-615	Sequence 615, App
20	37	33.9	742	6	US-10-995-561-618	Sequence 618, App
21	37	33.9	1448	6	US-10-485-517-212	Sequence 212, App
22	36	33.0	134	6	US-10-667-295-184	Sequence 184, App
23	36	33.0	245	6	US-10-131-826A-214	Sequence 214, App
24	36	33.0	306	6	US-10-793-626-382	Sequence 382, App
25	36	33.0	456	6	US-10-793-626-1190	Sequence 1190, Ap
26	36	33.0	1493	7	US-11-004-057-4	Sequence 4, Appli
27	36	33.0	2004	6	US-10-467-657-84	Sequence 84, Appl
28	36	33.0	2004	6	US-10-467-657-6322	Sequence 6322, Ap
29	35.5	32.6	371	6	US-10-467-657-1926	Sequence 1926, Ap
30	35.5	32.6	500	6	US-10-957-569-18	Sequence 18, Appl
31	35	32.1	67	6	US-10-467-657-5586	Sequence 5586, Ap
32	35	32.1	74	6	US-10-467-657-4748	Sequence 4748, Ap
33	35	32.1	205	7	US-11-093-746A-26	Sequence 26, Appl
34	35	32.1	235	6	US-10-793-626-948	Sequence 948, App
35	35	32.1	285	6	US-10-467-657-100	Sequence 100, App
36	35	32.1	297	6	US-10-467-657-7296	Sequence 7296, Ap
37	35	32.1	305	6	US-10-793-626-2062	Sequence 2062, Ap
38	35	32.1	353	7	US-11-060-029-23	Sequence 23, Appl
39	35	32.1	357	6	US-10-467-657-8540	Sequence 8540, Ap
40	35	32.1	366	7	US-11-000-463-410	Sequence 410, App
41	35	32.1	366	7	US-11-000-463-882	Sequence 882, App
42	35	32.1	386	6	US-10-858-730-236	Sequence 236, App
43	35	32.1	476	6	US-10-763-712A-50	Sequence 50, Appl
44	35	32.1	534	7	US-11-167-856-24	Sequence 24, Appl
45	35	32.1	856	6	US-10-467-657-8534	Sequence 8534, Ap

#### ALIGNMENTS

RESULT 1

US-10-467-657-4550

; Sequence 4550, Application US/10467657  
; Publication No. US20050260581A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON SpA  
; APPLICANT: FONTANA Maria Rita  
; APPLICANT: PIZZA Mariagrazia  
; APPLICANT: MASIGNANI Vega  
; APPLICANT: MONACI Elisabetta  
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/10/467,657  
; CURRENT FILING DATE: 2003-08-11  
; PRIOR APPLICATION NUMBER: GB-0103424.8  
; PRIOR FILING DATE: 2001-02-12  
; NUMBER OF SEQ ID NOS: 9218  
; SOFTWARE: SeqWin99, version 1.04  
; SEQ ID NO 4550  
; LENGTH: 436  
; TYPE: PRT  
; ORGANISM: Neisseria gonorrhoeae  
US-10-467-657-4550

Query Match 40.4%; Score 44; DB 6; Length 436;  
Best Local Similarity 39.1%; Pred. No. 6.8;  
Matches 9; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

Qy 1 AASSVSSASSRSYDYSRRNVRKN 23  
||:: |:| ::|: || |:|  
Db 332 AAAAKSNAGYKAYNQMRFRVKEN 354

Search completed: December 17, 2005, 00:17:54  
Job time : 7.16071 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2005, 23:33:32 ; Search time 23.8214 Seconds  
(without alignments)  
92.899 Million cell updates/sec

Title: US-10-789-494B-3  
Perfect score: 109  
Sequence: 1 AASSVSSASSRSYDYSRRNVRKN 23

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: , 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_80:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query Match	Length			
1	55	50.5	680	2	S52820	hypothetical prote
2	49	45.0	383	2	S39651	repB protein - Lac
3	47.5	43.6	674	2	S61181	hypothetical prote
4	47	43.1	766	2	G71437	probable resistanc
5	47	43.1	1041	2	C85189	disease resistance
6	47	43.1	1256	2	C71436	probable resistanc
7	46	42.2	386	2	T30320	replication protei
8	46	42.2	460	2	B45353	orf2 protein - app
9	46	42.2	684	2	T40319	hypothetical prote
10	45	41.3	94	2	A97140	phage related tran
11	45	41.3	381	2	T40341	hypothetical prote
12	45	41.3	486	2	I55449	gene PP2A protein
13	45	41.3	892	2	T09193	ataxin 7 - human
14	45	41.3	947	2	T03795	ornithine decarbox
15	45	41.3	1289	2	T00387	KIAA0622 protein -
16	45	41.3	1807	2	S03124	vitellogenin A2 pr
17	44.5	40.8	1677	2	T43021	vitellogenin precu
18	44	40.4	262	2	T14807	hypothetical prote
19	44	40.4	490	2	S53639	protein kinase clk
20	44	40.4	490	2	S70352	protein kinase clk
21	44	40.4	754	1	P1IV50	RNA-directed RNA p
22	44	40.4	1513	2	S45768	mitotic spindle pr
23	43	39.4	342	2	C84234	phosphate regulato
24	43	39.4	369	2	C88030	protein F46F5.10 [
25	43	39.4	382	2	T04260	hypothetical prote
26	43	39.4	405	2	C42094	bw3 protein - smut
27	43	39.4	602	2	S58336	probable membrane
28	43	39.4	695	2	T13648	mitosis initiation
29	43	39.4	1108	2	A96623	hypothetical prote
30	43	39.4	1383	2	C96723	hypothetical prote
31	43	39.4	1465	2	D86478	protein F15O4.11 [
32	42	38.5	90	2	S73849	MG220 homolog F10_
33	42	38.5	133	2	B30902	probable membrane



34	42	38.5	181	1	TXSPM	thioredoxin m prec
35	42	38.5	226	2	A86316	protein T10022.3 [
36	42	38.5	275	2	T48696	mRNA splicing fact
37	42	38.5	305	2	A56554	transcription fact
38	42	38.5	333	2	H96814	hypothetical prote
39	42	38.5	522	2	T28113	hypothetical prote
40	42	38.5	551	2	S66701	probable membrane
41	42	38.5	565	1	HMIVTN	hemagglutinin prec
42	42	38.5	605	2	S46833	hypothetical prote
43	42	38.5	629	2	JG0195	protein kinase DYR
44	42	38.5	923	2	S50482	hypothetical prote
45	42	38.5	1270	2	T22615	hypothetical prote

# ALIGNMENTS

## RESULT 1

S52820

hypothetical protein YPR007c - yeast (*Saccharomyces cerevisiae*)

N;Alternate names: hypothetical protein LPZ7c; hypothetical protein YP9723.07c

C;Species: *Saccharomyces cerevisiae*

C;Date: 19-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 09-Jul-2004

C;Accession: S52820; S59752

R;Pearson, D.; Bowman, S.

submitted to the EMBL Data Library, April 1995

A;Reference number: S52814

A;Accession: S52820

A;Molecule type: DNA

A;Residues: 1-680 <PEA>

A;Cross-references: UNIPROT:Q12188; UNIPARC:UPI000006B335; EMBL:Z48951;

NID:g762999; PID:g763006; MIPS:YPR007c

A;Experimental source: strain AB972

R;Wang, Y.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Hall, J.; Storms, R.K.; Vo, D.H.; Winnett, E.

submitted to the EMBL Data Library, July 1995

A;Description: The sequence of *Saccharomyces cerevisiae* chromosome XVI right arm.

A;Reference number: S59746

A;Accession: S59752

A;Molecule type: DNA

A;Residues: 1-680 <WAN>

A;Cross-references: UNIPARC:UPI000006B335; EMBL:U31900; NID:g1276597;

PID:g939741; MIPS:YPR007c

C;Genetics:

A;Gene: SGD:SPO69

A;Cross-references: SGD:S0006211

A;Map position: 16R

Query Match 50.5%; Score 55; DB 2; Length 680;

Best Local Similarity 47.6%; Pred. No. 2.1;

Matches 10; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

Qy 3 SSVSSASSRSYDYSRRNVRKN 23

|: ||:|:|:| |:: | |

Db 439 STQSSSSTRSHEYGRKSFRNN 459

Search completed: December 16, 2005, 23:53:14  
Job time : 25.8214 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2005, 23:33:02 ; Search time 147.036 Seconds  
(without alignments)  
110.362 Million cell updates/sec

Title: US-10-789-494B-3  
Perfect score: 109  
Sequence: 1 AASSVSSASSRSYDYSRRNVRKN 23

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_05.80:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result		%					
No.	Score	Query Match	Length	DB	ID	Description	
1	109	100.0	5263	1	FIBH_BOMMO	P05790 bombyx mori	
2	55	50.5	680	2	Q12188_YEAST	Q12188 saccharomyc	
3	50	45.9	271	1	RRFC_SPIOL	P82231 spinacia ol	
4	50	45.9	663	1	DUS8_MOUSE	O09112 mus musculu	
5	50	45.9	665	2	Q7TSZ9_MOUSE	Q7tsz9 mus musculu	
6	50	45.9	1059	2	Q9VW51_DROME	Q9vw51 drosophila	
7	50	45.9	1060	2	Q95VE6_DROME	Q95ve6 drosophila	
8	49	45.0	158	2	Q8GPL9_METEX	Q8gpl9 methylobact	
9	49	45.0	237	2	Q57W53_9TRYP	Q57w53 trypanosoma	
10	49	45.0	341	2	Q54NX5_DICDI	Q54nx5 dictyosteli	
11	49	45.0	351	2	Q6BL67_DEBHA	Q6bl67 debaryomyce	

12	49	45.0	383	2	Q07137_9LACT	Q07137 lactococcus
13	49	45.0	476	2	Q759N7_ASHGO	Q759n7 ashbya goss
14	49	45.0	820	2	Q4IP00_GIBZE	Q4ip00 gibberella
15	49	45.0	1362	2	Q611V7_CAEBR	Q611v7 caenorhabdi
16	49	45.0	1695	2	Q4P3I9_USTMA	Q4p3i9 ustilago ma
17	49	45.0	2303	2	Q4PF49_USTMA	Q4pf49 ustilago ma
18	48	44.0	472	2	O52491_AGRVI	O52491 agrobacteri
19	48	44.0	472	2	Q9X5L2_AGRVI	Q9x5l2 agrobacteri
20	48	44.0	654	2	Q4Z078_PLABE	Q4z078 plasmodium
21	48	44.0	679	1	UBC17_CAEL	Q11076 caenorhabdi
22	48	44.0	1030	2	Q8H037_ORYSA	Q8h037 oryza sativ
23	48	44.0	1089	2	Q4QBW2_LEIMA	Q4qbw2 leishmania
24	47.5	43.6	674	2	Q06629_YEAST	Q06629 saccharomyc
25	47	43.1	110	2	Q7TUE3_PROMP	Q7tue3 prochloroco
26	47	43.1	157	2	Q8GW40_ARATH	Q8gw40 arabidopsis
27	47	43.1	306	2	Q99JI2_MOUSE	Q99ji2 mus musculu
28	47	43.1	393	2	Q8BNS4_MOUSE	Q8bns4 mus musculu
29	47	43.1	434	2	Q4QAI2_LEIMA	Q4qai2 leishmania
30	47	43.1	482	2	Q8BSN1_MOUSE	Q8bsn1 mus musculu
31	47	43.1	494	2	Q7RDC8_PLAYO	Q7rdc8 plasmodium
32	47	43.1	1041	2	O23536_ARATH	O23536 arabidopsis
33	47	43.1	1135	2	Q8S4Q0_ARATH	Q8s4q0 arabidopsis
34	47	43.1	1147	2	Q8L843_ARATH	Q8l843 arabidopsis
35	47	43.1	1150	2	Q50PT4_ENTHI	Q50pt4 entamoeba h
36	47	43.1	1232	2	Q8IEF5_PLAF7	Q8ief5 plasmodium
37	47	43.1	1256	2	O23528_ARATH	O23528 arabidopsis
38	47	43.1	1267	2	Q54UP9_DICDI	Q54up9 dictyosteli
39	47	43.1	1526	2	Q7RFW3_PLAYO	Q7rfw3 plasmodium
40	46.5	42.7	634	2	Q4FXZ9_LEIMA	Q4fxz9 leishmania
41	46	42.2	265	2	Q5U926_9CALI	Q5u926 norovirus h
42	46	42.2	371	2	Q8BZ78_MOUSE	Q8bz78 mus musculu
43	46	42.2	386	2	O54428_9LACT	O54428 lactococcus
44	46	42.2	386	2	Q48610_9LACT	Q48610 lactococcus
45	46	42.2	386	2	Q9FB63_LACLA	Q9fb63 lactococcus

# ALIGNMENTS

## RESULT 1

FIBH\_BOMMO

ID FIBH\_BOMMO STANDARD; PRT; 5263 AA.

AC P05790; Q17220; Q26379;

DT 01-NOV-1988 (Rel. 09, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 13-SEP-2005 (Rel. 48, Last annotation update)

DE Fibroin heavy chain precursor (Fib-H) (H-fibroin).

GN Name=FIBH;

OS Bombyx mori (Silk moth).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;

OC Bombycidae; Bombyx.

OX NCBI\_TaxID=7091;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RX MEDLINE=20330362; PubMed=10871375; DOI=10.1093/nar/28.12.2413;

RA Zhou C.-Z., Confalonieri F., Medina N., Zivanovic Y., Esnault C.,

RA Yang T., Jacquet M., Janin J., Duguet M., Perasso R., Li Z.-G.;  
 RT "Fine organization of Bombyx mori fibroin heavy chain gene."  
 RL Nucleic Acids Res. 28:2413-2419(2000).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE OF 1-168.  
 RX MEDLINE=80045039; PubMed=498286; DOI=10.1016/0092-8674(79)90075-8;  
 RA Tsujimoto Y., Suzuki Y.;  
 RT "The DNA sequence of Bombyx mori fibroin gene including the 5'  
 RT flanking, mRNA coding, entire intervening and fibroin protein coding  
 RT regions.";  
 RL Cell 18:591-600(1979).  
 RN [3]  
 RP PARTIAL NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=79211211; PubMed=455439; DOI=10.1016/0092-8674(79)90018-7;  
 RA Tsujimoto Y., Suzuki Y.;  
 RT "Structural analysis of the fibroin gene at the 5' end and its  
 RT surrounding regions.";  
 RL Cell 16:425-436(1979).  
 RN [4]  
 RP PARTIAL NUCLEOTIDE SEQUENCE.  
 RC STRAIN=Kinshu X Showa;  
 RX MEDLINE=89094868; PubMed=3210244;  
 RA Mita K., Ichimura S., Zama M., James T.C.;  
 RT "Specific codon usage pattern and its implications on the secondary  
 RT structure of silk fibroin mRNA.";  
 RL J. Mol. Biol. 203:917-925(1988).  
 RN [5]  
 RP PARTIAL NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=94365842; PubMed=7916056; DOI=10.1007/BF00175878;  
 RA Mita K., Ichimura S., James T.C.;  
 RT "Highly repetitive structure and its organization of the silk fibroin  
 RT gene.";  
 RL J. Mol. Evol. 38:583-592(1994).  
 RN [6]  
 RP NUCLEOTIDE SEQUENCE OF 5179-5263, AND DISULFIDE BONDS.  
 RC STRAIN=J-139;  
 RX MEDLINE=99296390; PubMed=10366732; DOI=10.1016/S0167-4838(99)00088-6;  
 RA Tanaka K., Kajiyama N., Ishikura K., Waga S., Kikuchi A., Ohtomo K.,  
 RA Takagi T., Mizuno S.;  
 RT "Determination of the site of disulfide linkage between heavy and  
 RT light chains of silk fibroin produced by Bombyx mori.";  
 RL Biochim. Biophys. Acta 1432:92-103(1999).  
 RN [7]  
 RP PARTIAL NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=Daizo;  
 RX PubMed=15591204; DOI=10.1126/science.1102210;  
 RA Xia Q., Zhou Z., Lu C., Cheng D., Dai F., Li B., Zhao P., Zha X.,  
 RA Cheng T., Chai C., Pan G., Xu J., Liu C., Lin Y., Qian J., Hou Y.,  
 RA Wu Z., Li G., Pan M., Li C., Shen Y., Lan X., Yuan L., Li T., Xu H.,  
 RA Yang G., Wan Y., Zhu Y., Yu M., Shen W., Wu D., Xiang Z., Yu J.,  
 RA Wang J., Li R., Shi J., Li H., Li G., Su J., Wang X., Li G., Zhang Z.,  
 RA Wu Q., Li J., Zhang Q., Wei N., Xu J., Sun H., Dong L., Liu D.,  
 RA Zhao S., Zhao X., Meng Q., Lan F., Huang X., Li Y., Fang L., Li C.,  
 RA Li D., Sun Y., Zhang Z., Yang Z., Huang Y., Xi Y., Qi Q., He D.,  
 RA Huang H., Zhang X., Wang Z., Li W., Cao Y., Yu Y., Yu H., Li J.,  
 RA Ye J., Chen H., Zhou Y., Liu B., Wang J., Ye J., Ji H., Li S., Ni P.,  
 RA Zhang J., Zhang Y., Zheng H., Mao B., Wang W., Ye C., Li S., Wang J.,

RA Wong G.K.-S., Yang H.;  
 RT "A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*).";  
 RL Science 306:1937-1940(2004).  
 RN [8]  
 RP SUBUNIT.  
 RX MEDLINE=20568317; PubMed=10986287; DOI=10.1074/jbc.M006897200;  
 RA Inoue S., Tanaka K., Arisaka F., Kimura S., Ohtomo K., Mizuno S.;  
 RT "Silk fibroin of *Bombyx mori* is secreted, assembling a high molecular  
 RT mass elementary unit consisting of H-chain, L-chain, and p25, with a  
 RT 6:6:1 molar ratio.";  
 RL J. Biol. Chem. 275:40517-40528(2000).  
 CC !- FUNCTION: Core component of the silk filament; a strong, insoluble  
 CC and chemically inert fiber.  
 CC !- SUBUNIT: Silk fibroin elementary unit consists in a disulfide-  
 CC linked heavy and light chain and a p25 glycoprotein in molar  
 CC ratios of 6:6:1. This results in a complex of approximately 2.3  
 CC MDa.  
 CC !- TISSUE SPECIFICITY: Produced exclusively in the posterior (PSG)  
 CC section of silk glands, which are essentially modified salivary  
 CC glands.  
 CC !- DOMAIN: Composed of antiparallel beta sheets. The strands of the  
 CC beta sheets run parallel to the fiber axis. Long stretches of silk  
 CC fibroin are composed of microcrystalline arrays of (-Gly-Ser-Gly-  
 CC Ala-Gly-Ala-)n interrupted by regions containing bulkier residues.  
 CC The fiber is composed of microcrystalline arrays alternating with  
 CC amorphous regions.  
 CC !- PTM: The interchain disulfide bridge is essential for the  
 CC intracellular transport and secretion of fibroin.  
 CC -----  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use as long as its content is in no way modified and this statement is not  
 CC removed.  
 CC -----  
 DR EMBL; AF226688; AAF76983.1; -; Genomic\_DNA.  
 DR EMBL; V00094; CAA23432.1; -; Genomic\_DNA.  
 DR EMBL; V00097; CAA23433.1; -; Genomic\_DNA.  
 DR EMBL; S74439; AAB31861.1; -; mRNA.  
 DR EMBL; X13869; CAA32076.1; -; mRNA.  
 DR EMBL; M35378; AAA27839.1; -; mRNA.  
 DR EMBL; AB017362; BAA33147.1; -; Genomic\_DNA.  
 DR EMBL; CK538369; -; NOT\_ANNOTATED\_CDS; mRNA.  
 DR EMBL; AADK01000575; -; NOT\_ANNOTATED\_CDS; Genomic\_DNA.  
 DR PIR; S01844; S01844.  
 KW Repeat; Signal; Silk.  
 FT SIGNAL 1 21 Potential.  
 FT CHAIN 22 5263 Fibroin heavy chain.  
 FT REGION 149 5206 Highly repetitive.  
 FT DISULFID 5244 5244 Interchain (with C-190 in light chain).  
 FT DISULFID 5260 5263  
 FT CONFLICT 10 10 C -> V (in Ref. 2).  
 SQ SEQUENCE 5263 AA; 391593 MW; 8EE11D3A0A47440E CRC64;

Query Match 100.0%; Score 109; DB 1; Length 5263;  
 Best Local Similarity 100.0%; Pred. No. 5.3e-07;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 AASSVSSASSRSYDYSRRNVRKN 23
          |||||||||||||||||||||
Db     5221 AASSVSSASSRSYDYSRRNVRKN 5243
```

Search completed: December 16, 2005, 23:52:12

Job time : 150.036 secs

OM protein - protein search, using sw model

Run on: December 16, 2005, 23:08:17 ; Search time 125 Seconds  
(without alignments)  
70.301 Million cell updates/sec

Title: US-10-789-494B-7  
Perfect score: 108  
Sequence: 1 VETIVLEEDPYGHEDIYEED 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_21:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*  
9: geneseqp2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	%		Query		DB	ID	Description
	Score	Match	Length				
1	108	100.0	20	8	ADU51237	Adu51237 Gut silkw	
2	108	100.0	20	8	ADU51211	Adu51211 Silkworm	
3	108	100.0	120	8	ADU51171	Adu51171 Gut silkw	
4	108	100.0	2655	7	ADO59401	Ado59401 Antheraea	
5	50	46.3	1100	5	ABB50120	Abb50120 Listeria	
6	49	45.4	299	9	ABM93718	Abm93718 M. xanthu	
7	49	45.4	325	3	AAB07672	Aab07672 Amino aci	
8	49	45.4	325	8	ADU66520	Adu66520 TaD, memb	

9	49	45.4	325	9	ADZ44577	Adz44577	Myxococcu
10	49	45.4	640	6	ABP71659	Abp71659	A. cellul
11	49	45.4	640	7	ADF75899	Adf75899	Catalytic
12	49	45.4	1121	6	ABP71656	Abp71656	A. cellul
13	49	45.4	1121	7	ADF75895	Adf75895	Acidother
14	48.5	44.9	201	7	ADM25426	Adm25426	Hyperther
15	48	44.4	55	5	ABG80758	Abg80758	C-type le
16	48	44.4	159	5	ABB56404	Abb56404	Human col
17	48	44.4	277	2	AAy25518	Aay25518	Human col
18	48	44.4	277	2	AAy41698	Aay41698	Human PRO
19	48	44.4	277	3	AAB44254	Aab44254	Human PRO
20	48	44.4	277	4	AAU29073	Aau29073	Human PRO
21	48	44.4	277	6	ABU58449	Abu58449	Human PRO
22	48	44.4	277	6	ABU87997	Abu87997	Novel hum
23	48	44.4	277	6	ABU84312	Abu84312	Human sec
24	48	44.4	277	6	ABR66186	Abr66186	Human sec
25	48	44.4	277	6	ABR65576	Abr65576	Human sec
26	48	44.4	277	6	ABU99516	Abu99516	Human sec
27	48	44.4	277	6	ABU82755	Abu82755	Human PRO
28	48	44.4	277	6	ABU89876	Abu89876	Novel hum
29	48	44.4	277	6	ABR68125	Abr68125	Human sec
30	48	44.4	277	6	ABU96178	Abu96178	Novel hum
31	48	44.4	277	6	ABU92609	Abu92609	Human sec
32	48	44.4	277	6	ABO08686	Abo08686	Human sec
33	48	44.4	277	6	ABO02738	Abo02738	Human sec
34	48	44.4	277	6	ABR74892	Abr74892	Human sec
35	48	44.4	277	6	ABR94654	Abr94654	Human sec
36	48	44.4	277	6	ABO25200	Abo25200	Novel hum
37	48	44.4	277	6	ABU85627	Abu85627	Human PRO
38	48	44.4	277	6	ABU98787	Abu98787	Novel hum
39	48	44.4	277	6	ABU98002	Abu98002	Novel hum
40	48	44.4	277	6	ABU91708	Abu91708	Novel hum
41	48	44.4	277	6	ABU72206	Abu72206	Novel hum
42	48	44.4	277	6	ABU89401	Abu89401	Human PRO
43	48	44.4	277	6	ABU86242	Abu86242	Human sec
44	48	44.4	277	6	ABU67455	Abu67455	Human sec
45	48	44.4	277	6	ABU80483	Abu80483	Human PRO

#### ALIGNMENTS

##### RESULT 1

ADU51237

ID ADU51237 standard; peptide; 20 AA.

XX

AC ADU51237;

XX

DT 24-FEB-2005 (first entry)

XX

DE Gut silkworm fibroin peptide fragment 39.

XX

KW vulnerary; cell proliferation; wound healing; cell adhesion; cosmetics;  
KW cell culture; fibroin.

XX

OS Bombycoidea.

XX



PN JP2004339189-A.  
 XX  
 PD 02-DEC-2004.  
 XX  
 PF 04-DEC-2003; 2003JP-00406608.  
 XX  
 PR 28-FEB-2003; 2003JP-00055048.  
 XX  
 PA (DOKU-) DOKURITSU GYOSEI HOJIN NOGYO SEIBUTSU SH.  
 PA (TSUB/) TSUBOUCHI K.  
 XX  
 DR WPI; 2004-827614/82.  
 XX  
 PT New peptide having excellent cell growth promoting activity, for use as a  
 PT cell growth promoter, cell adhesion agent, wound healing-promoting agent,  
 PT cosmetic and cell culture base material.  
 XX  
 PS Example 3; Page; 27pp; Japanese.  
 XX  
 CC The invention relates to a novel peptide having excellent cell growth  
 CC promoting activity. The peptide of the invention demonstrates vulnerary  
 CC activity and may be utilised as a cell growth promoter, cell adhesion  
 CC agent, wound healing-promoting agent or cosmetic and cell culture base  
 CC material. The current sequence is that of a gut silkworm fibroin peptide  
 CC fragment of the invention which is described as being amorphous.  
 XX  
 SQ Sequence 20 AA;

Query Match 100.0%; Score 108; DB 8; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 3.6e-10;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VETIVLEEDPYGHEDIYEED 20  
 |||||  
 Db 1 VETIVLEEDPYGHEDIYEED 20

#### RESULT 4

ADO59401

ID ADO59401 standard; protein; 2655 AA.

XX

AC ADO59401;

XX

DT 15-JUL-2004 (first entry)

XX

DE Antheraea yamamai fibroin.

XX

KW Japanese oak silkmoth; fibroin; silkworm; raw silk; thread.

XX

OS Antheraea yamamai.

XX

PN KR2002094304-A.

XX

PD 18-DEC-2002.

XX

PF 08-JUN-2001; 2001KR-00031954.

XX  
 PR 08-JUN-2001; 2001KR-00031954.  
 XX  
 PA (RURA-) RURAL DEV ADMINISTRATION.  
 XX  
 PI Hwang JS, Jin BR, Kang SU, Kim GY, Kim YS, Koo TW, Lee JS;  
 PI Lee SM, Seo DS, Yoon EY;  
 XX  
 DR WPI; 2003-399638/38.  
 DR N-PSDB; ADO59400.  
 XX  
 PT Nucleotide sequence encoding fibroin useful for transforming silkworms  
 PT for producing raw silk thread comprises a gene isolated from Antheraea  
 PT yamamai and its promoter.  
 XX  
 PS Example 1; SEQ ID NO 2; 19pp; Korean.  
 XX  
 CC This invention describes a novel nucleic acid isolated from Antheraea  
 CC yamamai (Japanese oak silkworm) which encodes fibroin. The nucleic acid  
 CC is useful for producing transformed silkworms capable of producing raw  
 CC silk-thread.  
 XX  
 SQ Sequence 2655 AA;

Query Match 100.0%; Score 108; DB 7; Length 2655;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VETIVLEEDPYGHEDIYEED 20  
 |||||  
 Db 61 VETIVLEEDPYGHEDIYEED 80

RESULT 5  
 ABB50120  
 ID ABB50120 standard; protein; 1100 AA.  
 XX  
 AC ABB50120;  
 XX  
 DT 05-FEB-2002 (first entry)  
 XX  
 DE Listeria monocytogenes protein #2824.  
 XX  
 KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 KW vitamin B12; bacterial infection; disease.  
 XX  
 OS Listeria monocytogenes.  
 XX  
 PN WO200177335-A2.  
 XX  
 PD 18-OCT-2001.  
 XX  
 PF 11-APR-2001; 2001WO-FR001118.  
 XX  
 PR 11-APR-2000; 2000FR-00004629.  
 XX  
 PA (INSP ) INST PASTEUR.

XX  
 PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;  
 PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;  
 PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;  
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;  
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
 PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;  
 PI Rose M, Voss H;  
 XX  
 DR WPI; 2002-010914/01.  
 XX  
 PT Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment  
 PT and prevention of *Listeria* and related bacterial infections, and related  
 PT polypeptides.  
 XX  
 PS Claim 6; SEQ ID NO 2825; 192pp; French.  
 XX  
 CC The present invention relates to the genome sequence of *Listeria*  
 CC *monocytogenes* EGD-e (see ABA03041). The genome sequence and fragments of  
 CC it are useful for selecting probes and primers for detecting genes in *L.*  
 CC *monocytogenes* and related organisms, and for studying genetic  
 CC polymorphisms and other genomes. The present sequence is a protein  
 CC encoded by the genome sequence of the present invention. Proteins  
 CC expressed from the genome sequence are useful for raising specific  
 CC antibodies, identification of *L. monocytogenes* and related organisms, and  
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
 CC B12. The genome sequence and proteins encoded by it are also useful for  
 CC selecting compounds that regulate gene expression and cell replication  
 CC and modulate *L. monocytogenes*-related diseases. In addition, the genome  
 CC sequence and proteins encoded by it are useful in pharmaceutical and  
 CC vaccines compositions for the treatment or prevention of infections by *L.*  
 CC *monocytogenes* and related organisms. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 1100 AA;

Query Match 46.3%; Score 50; DB 5; Length 1100;  
 Best Local Similarity 47.1%; Pred. No. 97;  
 Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 3 TIVLEEDPYGHEDIYEE 19  
 | : ||:||||: ||:  
 Db 242 TAHISHDPFGHENRYEQ 258

Search completed: December 16, 2005, 23:46:10  
 Job time : 127 secs

OM protein - protein search, using sw model

Run on: December 16, 2005, 23:40:43 ; Search time 30.3571 Seconds  
(without alignments)  
54.469 Million cell updates/sec

Title: US-10-789-494B-7  
Perfect score: 108  
Sequence: 1 VETIVLEEDPYGHEDIYEED 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/6\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/H\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/RE\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	49	45.4	299	2	US-09-902-540-12917	Sequence 12917, A
2	49	45.4	325	2	US-09-710-262E-9	Sequence 9, Appli
3	48	44.4	277	2	US-09-600-932-2	Sequence 2, Appli
4	48	44.4	277	2	US-09-999-833A-97	Sequence 97, Appl
5	48	44.4	277	2	US-10-020-445A-97	Sequence 97, Appl
6	47	43.5	562	2	US-09-949-002-426	Sequence 426, App
7	47	43.5	636	2	US-09-564-805-237	Sequence 237, App
8	47	43.5	726	2	US-09-187-330-57	Sequence 57, Appl
9	47	43.5	781	2	US-09-187-330-32	Sequence 32, Appl
10	47	43.5	787	2	US-09-187-330-31	Sequence 31, Appl
11	47	43.5	800	2	US-09-187-330-41	Sequence 41, Appl
12	47	43.5	806	2	US-09-187-330-3	Sequence 3, Appli
13	47	43.5	828	2	US-09-187-330-55	Sequence 55, Appl
14	47	43.5	874	2	US-09-187-330-59	Sequence 59, Appl
15	47	43.5	937	1	US-08-253-155A-31	Sequence 31, Appl
16	47	43.5	937	2	US-09-538-092-1092	Sequence 1092, Ap

17	47	43.5	937	2	US-09-949-002-374	Sequence 374, App
18	47	43.5	959	2	US-09-538-092-1091	Sequence 1091, Ap
19	47	43.5	1000	2	US-09-187-330-1	Sequence 1, Appli
20	47	43.5	1102	2	US-09-364-609-8	Sequence 8, Appli
21	46.5	43.1	289	2	US-09-902-540-10736	Sequence 10736, A
22	46	42.6	276	2	US-09-902-540-16693	Sequence 16693, A
23	45	41.7	512	2	US-09-949-016-10977	Sequence 10977, A
24	45	41.7	513	1	US-08-357-533A-11	Sequence 11, Appl
25	45	41.7	513	1	US-08-459-009-11	Sequence 11, Appl
26	45	41.7	513	2	US-08-459-951-11	Sequence 11, Appl
27	45	41.7	799	2	US-08-909-954-4	Sequence 4, Appli
28	45	41.7	804	2	US-08-909-954-2	Sequence 2, Appli
29	44	40.7	331	2	US-09-248-796A-18164	Sequence 18164, A
30	44	40.7	357	1	US-08-978-182-4	Sequence 4, Appli
31	44	40.7	357	1	US-09-205-681-4	Sequence 4, Appli
32	44	40.7	373	2	US-09-248-796A-20443	Sequence 20443, A
33	44	40.7	850	2	US-09-583-110-4394	Sequence 4394, Ap
34	44	40.7	861	2	US-09-107-433-2860	Sequence 2860, Ap
35	43	39.8	238	2	US-09-489-039A-9218	Sequence 9218, Ap
36	43	39.8	302	2	US-09-438-185A-457	Sequence 457, App
37	43	39.8	324	2	US-09-248-796A-15787	Sequence 15787, A
38	43	39.8	342	2	US-09-248-796A-14202	Sequence 14202, A
39	43	39.8	357	2	US-09-198-452A-487	Sequence 487, App
40	43	39.8	455	2	US-09-543-681A-8288	Sequence 8288, Ap
41	43	39.8	536	1	US-08-357-533A-12	Sequence 12, Appl
42	43	39.8	536	1	US-08-459-009-12	Sequence 12, Appl
43	43	39.8	536	2	US-08-459-951-12	Sequence 12, Appl
44	43	39.8	536	2	US-09-267-963D-35	Sequence 35, Appl
45	43	39.8	817	1	US-08-381-931B-2	Sequence 2, Appli

#### ALIGNMENTS

##### RESULT 1

```

US-09-902-540-12917
; Sequence 12917, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 12917
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-12917

```

Query Match 45.4%; Score 49; DB 2; Length 299;

Best Local Similarity 53.8%; Pred. No. 7.2;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 7 EEDPYGHEDIYEE 19  
| | :| | | :| | :  
Db 95 EADAFGHEELYEQ 107

Search completed: December 17, 2005, 00:12:07  
Job time : 30.3571 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2005, 23:52:28 ; Search time 105.714 Seconds  
(without alignments)  
79.049 Million cell updates/sec

Title: US-10-789-494B-7  
Perfect score: 108  
Sequence: 1 VETIVLEEDPYGHEDIYEED 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA\_Main:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result	%	Query				
No.	Score	Match Length	DB	ID		Description

1	108	100.0	20	5	US-10-789-494B-7	Sequence 7, Appli
2	108	100.0	20	5	US-10-789-494B-67	Sequence 67, Appl
3	108	100.0	120	5	US-10-789-494B-23	Sequence 23, Appl
4	53	49.1	127	4	US-10-424-599-203223	Sequence 203223,
5	51	47.2	113	4	US-10-424-599-200988	Sequence 200988,
6	49	45.4	325	5	US-10-848-111-9	Sequence 9, Appli
7	49	45.4	554	4	US-10-424-599-180031	Sequence 180031,
8	48	44.4	159	4	US-10-258-105-14	Sequence 14, Appl
9	48	44.4	277	3	US-09-978-295A-97	Sequence 97, Appl
10	48	44.4	277	3	US-09-978-697-97	Sequence 97, Appl
11	48	44.4	277	3	US-09-978-192A-97	Sequence 97, Appl
12	48	44.4	277	3	US-09-999-832A-97	Sequence 97, Appl
13	48	44.4	277	3	US-09-978-189-97	Sequence 97, Appl
14	48	44.4	277	3	US-09-978-608A-97	Sequence 97, Appl
15	48	44.4	277	3	US-09-978-585A-97	Sequence 97, Appl
16	48	44.4	277	3	US-09-978-191A-97	Sequence 97, Appl
17	48	44.4	277	3	US-09-978-403A-97	Sequence 97, Appl
18	48	44.4	277	3	US-09-978-564A-97	Sequence 97, Appl
19	48	44.4	277	3	US-09-999-833A-97	Sequence 97, Appl
20	48	44.4	277	3	US-09-981-915A-97	Sequence 97, Appl
21	48	44.4	277	3	US-09-978-824-97	Sequence 97, Appl
22	48	44.4	277	3	US-09-918-585A-97	Sequence 97, Appl
23	48	44.4	277	3	US-09-999-834A-97	Sequence 97, Appl
24	48	44.4	277	3	US-09-978-423A-97	Sequence 97, Appl
25	48	44.4	277	3	US-09-978-193A-97	Sequence 97, Appl
26	48	44.4	277	3	US-09-999-830A-97	Sequence 97, Appl
27	48	44.4	277	3	US-09-978-757A-97	Sequence 97, Appl
28	48	44.4	277	3	US-09-978-187B-97	Sequence 97, Appl
29	48	44.4	277	3	US-09-978-643A-97	Sequence 97, Appl
30	48	44.4	277	3	US-09-978-375A-97	Sequence 97, Appl
31	48	44.4	277	3	US-09-978-298A-97	Sequence 97, Appl
32	48	44.4	277	3	US-09-978-188A-97	Sequence 97, Appl
33	48	44.4	277	3	US-09-978-681A-97	Sequence 97, Appl
34	48	44.4	277	3	US-09-978-194A-97	Sequence 97, Appl
35	48	44.4	277	3	US-09-999-829A-97	Sequence 97, Appl
36	48	44.4	277	3	US-09-978-299A-97	Sequence 97, Appl
37	48	44.4	277	3	US-09-978-544A-97	Sequence 97, Appl
38	48	44.4	277	3	US-09-978-665A-97	Sequence 97, Appl
39	48	44.4	277	3	US-09-978-802A-97	Sequence 97, Appl
40	48	44.4	277	3	US-09-999-831A-97	Sequence 97, Appl
41	48	44.4	277	3	US-09-978-824-97	Sequence 97, Appl
42	48	44.4	277	4	US-10-052-586-100	Sequence 100, App
43	48	44.4	277	4	US-10-174-590-100	Sequence 100, App
44	48	44.4	277	4	US-10-176-758-100	Sequence 100, App
45	48	44.4	277	4	US-10-175-737-100	Sequence 100, App

#### ALIGNMENTS

RESULT 1

US-10-789-494B-7

; Sequence 7, Application US/10789494B

; Publication No. US20050143296A1

; GENERAL INFORMATION:

; APPLICANT: TSUBOUCHI, Kozo

```
; APPLICANT: YAMADA, Hiromi
; TITLE OF INVENTION: EXTRACTION AND UTILIZATION OF CELL
; TITLE OF INVENTION: GROWTH-PROMOTING PEPTIDES FROM SILK PROTEIN
; FILE REFERENCE: OPS 635
; CURRENT APPLICATION NUMBER: US/10/789,494B
; CURRENT FILING DATE: 2004-02-27
; PRIOR APPLICATION NUMBER: JP 2003-55048
; PRIOR FILING DATE: 2003-02-28
; NUMBER OF SEQ ID NOS: 85
; SEQ ID NO 7
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Antheraea yamamai
US-10-789-494B-7
```

```
Query Match          100.0%; Score 108; DB 5; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 VETIVLEEDPYGHEDIYEED 20
          |||||
Db      1 VETIVLEEDPYGHEDIYEED 20
```

#### RESULT 4

```
US-10-424-599-203223
; Sequence 203223, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated
With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 203223
; LENGTH: 127
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_25535C.1.pep
US-10-424-599-203223
```

```
Query Match          49.1%; Score 53; DB 4; Length 127;
Best Local Similarity 50.0%; Pred. No. 4.2;
Matches 10; Conservative 5; Mismatches 3; Indels 2; Gaps 1;
```

```
QY      3 TIVLEED--PYGHEDIYEED 20
          ::| || | ||: :|||
Db     108 SVILNEDSRPCGHKCLYEED 127
```



Search completed: December 17, 2005, 00:17:33  
Job time : 106.714 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2005, 23:53:24 ; Search time 5.35714 Seconds  
(without alignments)  
25.151 Million cell updates/sec

Title: US-10-789-494B-7  
Perfect score: 108  
Sequence: 1 VETIVLEEDPYGHEDIYEED 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 51470 seqs, 6736768 residues

Total number of hits satisfying chosen parameters: 51470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA\_New:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result		%					
No.	Score	Query	Match	Length	DB	ID	Description
1	45	41.7	98	6	US-10-467-657-3490	Sequence 3490, Ap	
2	43	39.8	963	6	US-10-467-962B-2	Sequence 2, Appli	
3	42	38.9	459	7	US-11-186-284-12	Sequence 12, Appl	
4	42	38.9	2647	6	US-10-821-234-1303	Sequence 1303, Ap	
5	40	37.0	21	6	US-10-939-890-584	Sequence 584, App	

6	40	37.0	331	6	US-10-995-561-977	Sequence 977, App
7	40	37.0	331	6	US-10-995-561-978	Sequence 978, App
8	40	37.0	3655	7	US-11-075-185-5	Sequence 5, Appli
9	39.5	36.6	450	7	US-11-074-176-212	Sequence 212, App
10	39	36.1	651	7	US-11-172-145-10	Sequence 10, Appl
11	39	36.1	1116	6	US-10-485-517-238	Sequence 238, App
12	39	36.1	1117	6	US-10-485-517-206	Sequence 206, App
13	39	36.1	1236	7	US-11-115-086-4	Sequence 4, Appli
14	38	35.2	414	6	US-10-467-657-2432	Sequence 2432, Ap
15	38	35.2	564	7	US-11-186-284-199	Sequence 199, App
16	37	34.3	216	6	US-10-793-626-2624	Sequence 2624, Ap
17	37	34.3	307	6	US-10-793-626-684	Sequence 684, App
18	37	34.3	457	6	US-10-763-712A-61	Sequence 61, Appl
19	37	34.3	1065	6	US-10-793-626-1212	Sequence 1212, Ap
20	37	34.3	1152	7	US-11-080-026-4	Sequence 4, Appli
21	36.5	33.8	126	6	US-10-821-234-1318	Sequence 1318, Ap
22	36	33.3	265	6	US-10-793-626-2422	Sequence 2422, Ap
23	36	33.3	272	6	US-10-793-626-2290	Sequence 2290, Ap
24	36	33.3	286	6	US-10-793-626-1020	Sequence 1020, Ap
25	36	33.3	356	6	US-10-793-626-2054	Sequence 2054, Ap
26	36	33.3	388	6	US-10-467-657-786	Sequence 786, App
27	36	33.3	402	6	US-10-485-517-422	Sequence 422, App
28	36	33.3	502	6	US-10-131-826A-548	Sequence 548, App
29	36	33.3	502	6	US-10-689-742-148	Sequence 148, App
30	36	33.3	505	6	US-10-467-657-5148	Sequence 5148, Ap
31	36	33.3	615	7	US-11-172-145-6	Sequence 6, Appli
32	36	33.3	617	7	US-11-172-145-8	Sequence 8, Appli
33	36	33.3	695	6	US-10-363-924-2	Sequence 2, Appli
34	36	33.3	892	6	US-10-507-275-3	Sequence 3, Appli
35	36	33.3	903	7	US-11-057-058-65	Sequence 65, Appl
36	36	33.3	904	6	US-10-507-275-5	Sequence 5, Appli
37	36	33.3	904	7	US-11-087-227-12	Sequence 12, Appl
38	36	33.3	1442	6	US-10-793-626-2052	Sequence 2052, Ap
39	35.5	32.9	635	6	US-10-821-234-927	Sequence 927, App
40	35.5	32.9	2261	6	US-10-995-561-600	Sequence 600, App
41	35	32.4	154	6	US-10-467-657-158	Sequence 158, App
42	35	32.4	154	6	US-10-467-657-6520	Sequence 6520, Ap
43	35	32.4	154	6	US-10-467-657-7634	Sequence 7634, Ap
44	35	32.4	182	7	US-11-000-463-717	Sequence 717, App
45	35	32.4	219	6	US-10-467-657-4956	Sequence 4956, Ap

# ALIGNMENTS

## RESULT 1

US-10-467-657-3490

; Sequence 3490, Application US/10467657

; Publication No. US20050260581A1

; GENERAL INFORMATION:

; APPLICANT: CHIRON SpA

; APPLICANT: FONTANA Maria Rita

; APPLICANT: PIZZA Mariagrazia

; APPLICANT: MASIGNANI Vega

; APPLICANT: MONACI Elisabetta

; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS

; FILE REFERENCE:

; CURRENT APPLICATION NUMBER: US/10/467,657  
; CURRENT FILING DATE: 2003-08-11  
; PRIOR APPLICATION NUMBER: GB-0103424.8  
; PRIOR FILING DATE: 2001-02-12  
; NUMBER OF SEQ ID NOS: 9218  
; SOFTWARE: SeqWin99, version 1.04  
; SEQ ID NO 3490  
; LENGTH: 98  
; TYPE: PRT  
; ORGANISM: Neisseria gonorrhoeae  
US-10-467-657-3490

Query Match 41.7%; Score 45; DB 6; Length 98;  
Best Local Similarity 53.8%; Pred. No. 1.4;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 7 EEDPYGHEDIYEE 19  
|:| | | :| |  
Db 75 EDDPYVHAGVYSE 87

Search completed: December 17, 2005, 00:17:54  
Job time : 5.35714 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2005, 23:33:32 ; Search time 20.7143 Seconds  
(without alignments)  
92.899 Million cell updates/sec

Title: US-10-789-494B-7  
Perfect score: 108  
Sequence: 1 VETIVLEEDPYGHEDIYEED 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_80:\*  
1: pirl:\*  
2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	108	100.0	2639	2	T31328	fibroin - Chinese
2	50	46.3	1100	2	AG1097	alpha-xylosidase a
3	47	43.5	221	2	A84638	hypothetical prote
4	47	43.5	636	2	F69027	cleavage and polya
5	47	43.5	937	1	S42366	endopeptidase La h
6	47	43.5	962	1	S57342	endopeptidase La h
7	47	43.5	1005	2	T12546	hypothetical prote
8	47	43.5	1090	2	S59077	cellulose 1,4-beta
9	46	42.6	68	2	S27148	gene 30.4 protein
10	46	42.6	216	2	E89848	conserved hypothet
11	45.5	42.1	288	2	AH1399	ABC transporter (A
12	45.5	42.1	288	2	AG1775	ABC transporter (A
13	45.5	42.1	1165	2	A70423	valine-tRNA ligase
14	45	41.7	92	2	E81209	conserved hypothet
15	45	41.7	96	2	E81786	conserved hypothet
16	45	41.7	129	2	F72425	2-amino-4-hydroxy-
17	45	41.7	286	1	JQ2154	frxC protein - Ple
18	45	41.7	288	2	AF2440	protochlorophyllid
19	45	41.7	314	2	F84044	malate dehydrogena
20	45	41.7	382	2	B49193	type II activin re
21	45	41.7	429	2	AB3550	NAD(P) transhydrog
22	45	41.7	504	2	B40829	activin receptor i
23	45	41.7	512	2	D40829	activin receptor i
24	45	41.7	512	2	I37134	activin type II re
25	45	41.7	513	2	JQ1484	activin receptor p
26	45	41.7	614	2	JH0500	zinc finger protei
27	45	41.7	726	2	D97012	probable processiv
28	45	41.7	2206	2	G71611	hypothetical prote
29	44	40.7	300	2	A81418	pseudouridylate sy
30	44	40.7	357	2	S43278	cell division cont
31	44	40.7	376	2	A71175	probable dehydroge
32	44	40.7	407	2	F72772	probable valine-py
33	44	40.7	410	2	E75208	probable valine-py
34	44	40.7	539	2	F75497	probable arginine
35	44	40.7	695	2	T40451	n-terminal acetylt
36	44	40.7	850	2	F95260	ABC transporter, p
37	44	40.7	850	2	A98126	conserved hypothet
38	44	40.7	973	2	T35238	probable secreted
39	43.5	40.3	526	1	G71081	probable helicase
40	43.5	40.3	789	2	T38423	hypothetical prote
41	43.5	40.3	1684	2	T02367	hypothetical prote
42	43	39.8	89	2	T27015	hypothetical prote
43	43	39.8	98	2	A70301	ribosomal protein
44	43	39.8	114	1	E64423	ychN protein homol
45	43	39.8	142	2	A97409	hypothetical prote

# ALIGNMENTS

## RESULT 1

T31328

fibroin - Chinese oak silkworm

C;Species: *Antheraea pernyi* (Chinese oak silkworm)

C;Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004

C;Accession: T31328

R;Sezutsu, H.; Tamura, T.; Yukuhiro, K.

submitted to the EMBL Data Library, August 1998

A;Description: Characterization of the full length fibroin gene of a wild silkworm, *Antheraea pernyi*.

A;Reference number: Z20995

A;Accession: T31328

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-2639 <SEZ>

A;Cross-references: UNIPROT:O76786; UNIPARC:UPI0000078D8E; EMBL:AF083334;

NID:g3450882; PID:g3450883; PIDN:AAC32606.1

C;Genetics:

A;Introns: 14/3

Query Match 100.0%; Score 108; DB 2; Length 2639;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VETIVLEEDPYGHEDIYEED 20  
|||  
Db 61 VETIVLEEDPYGHEDIYEED 80

## RESULT 2

AG1097

alpha-xylosidase and alpha-glucosidase homolog lmo0182 [imported] - *Listeria monocytogenes* (strain EGD-e)

C;Species: *Listeria monocytogenes*

C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004

C;Accession: AG1097

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, H.; Brandt, P.; Chakraborty, T.; Charbit, A.; Chetouani, F.; Couve, E.; de Daruvar, A.; Dehoux, P.; Domann, E.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Garcia-Del Portillo, F.; Garrido, P.; Gautier, L.; Goebel, W.; Gomez-Lopez, N.; Hain, T.; Hauf, J.; Jackson, D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001

A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mata Vicente, J.; Ng, E.; Nordsiek, G.; Novella, S.; de Pablos, B.; Perez-Diaz, J.C.; Remmel, B.; Rose, M.; Rusniok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, J.; Cossart, P.

A;Title: Comparative genomics of *Listeria* species.

A;Reference number: AB1077; MUID:21537279; PMID:11679669

A;Accession: AG1097

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1100 <GLA>  
A;Cross-references: UNIPROT:Q8YAE8; UNIPARC:UPI0000055794; GB:NC\_003210;  
PIDN:CAC98397.1; PID:g16409539; GSPDB:GN00177  
A;Experimental source: strain EGD-e  
C;Genetics:  
A;Gene: lmo0182

Query Match 46.3%; Score 50; DB 2; Length 1100;  
Best Local Similarity 47.1%; Pred. No. 28;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 3 TIVLEEDPYGHEDIYEE 19  
| : ||:|||: ||:  
Db 242 TAHISHDPFGHENRYEQ 258

Search completed: December 16, 2005, 23:53:16  
Job time : 22.7143 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2005, 23:33:02 ; Search time 127.857 Seconds  
(without alignments)  
110.362 Million cell updates/sec

Title: US-10-789-494B-7  
Perfect score: 108  
Sequence: 1 VETIVLEEDPYGHEDIYEED 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_05.80:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result		%					
No.	Score	Query Match	Length	DB	ID	Description	
1	108	100.0	507	2	Q8ISB3_9NEOP	Q8isb3	antheraea m
2	108	100.0	2639	2	O76786_ANTPE	O76786	antheraea p
3	108	100.0	2655	2	Q964F4_ANTYA	Q964f4	antheraea y
4	66	61.1	1468	2	Q9GUB5_GALME	Q9gub5	galleria me
5	58	53.7	509	2	Q7Z010_PLOIN	Q7z010	plodia inte
6	58	53.7	735	2	Q7Z009_9NEOP	Q7z009	anagasta ku
7	56	51.9	1065	1	DPO3A_STAAR	Q6gg04	staphylococ
8	53	49.1	275	2	Q7NI14_GLOVI	Q7ni14	gloeobacter
9	50	46.3	358	2	Q8I2X6_PLAF7	Q8i2x6	plasmodium
10	50	46.3	447	2	Q8PWT5_METMA	Q8pwt5	methanosarc
11	50	46.3	1100	2	Q724N1_LISMF	Q724n1	listeria mo
12	50	46.3	1100	2	Q8YAE8_LISMO	Q8yae8	listeria mo
13	49	45.4	321	2	Q4JZ47_STRPN	Q4jz47	streptococc
14	49	45.4	321	2	Q4K056_STRPN	Q4k056	streptococc
15	49	45.4	325	2	Q9XB05_MYXXA	Q9xb05	myxococcus
16	49	45.4	359	2	Q74ZB3_ASHGO	Q74zb3	ashbya goss
17	49	45.4	569	2	Q869X0_DICDI	Q869x0	dictyosteli
18	49	45.4	637	2	Q5CC74_9CUCU	Q5cc74	otiorhynchu
19	49	45.4	721	2	Q609H0_METCA	Q609h0	methylococc
20	49	45.4	2482	2	Q7YYZ4_CRYPV	Q7yyz4	cryptospori
21	49	45.4	4605	2	Q8I1R3_PLAF7	Q8ilr3	plasmodium
22	48.5	44.9	201	2	Q8TZA3_METKA	Q8tza3	methanopyru
23	48.5	44.9	569	2	Q6BXH0_DEBHA	Q6bxh0	debaryomyce
24	48.5	44.9	619	2	Q8IEN9_PLAF7	Q8ien9	plasmodium
25	48	44.4	209	2	Q60ZN9_CAEBR	Q60zn9	caenorhabdi
26	48	44.4	277	2	Q6UW19_HUMAN	Q6uwl9	homo sapien
27	48	44.4	277	2	Q9Y6Z7_HUMAN	Q9y6z7	homo sapien
28	48	44.4	864	2	Q8IL99_PLAF7	Q8il99	plasmodium
29	48	44.4	1729	2	Q61AU4_CAEBR	Q61au4	caenorhabdi
30	47	43.5	53	2	Q6LE96_CELFI	Q6le96	cellulomona
31	47	43.5	121	1	Y008_METKA	P94948	methanopyru
32	47	43.5	315	1	Y034_METMP	Q6m183	methanococc
33	47	43.5	334	2	Q4LGU1_9BURK	Q4lgul	burkholderi
34	47	43.5	382	2	Q89IA2_BRAJA	Q89ia2	bradyrhizob
35	47	43.5	517	2	Q9BSN5_HUMAN	Q9bsn5	homo sapien
36	47	43.5	540	2	Q6N5W7_RHOPA	Q6n5w7	rhodopseudo
37	47	43.5	636	2	O27271_METTH	O27271	methanobact
38	47	43.5	727	2	O65986_CLOCL	O65986	clostridium
39	47	43.5	737	2	Q4IMY1_GIBZE	Q4imy1	gibberella
40	47	43.5	753	2	Q8J1E3_9FUNG	Q8j1e3	piromyces s
41	47	43.5	823	1	ADNP_RAT	Q9jkl8	rattus norv
42	47	43.5	828	1	ADNP_MOUSE	Q9z103	mus musculu
43	47	43.5	895	2	Q8N8K8_HUMAN	Q8n8k8	homo sapien
44	47	43.5	919	2	Q91BB5_NPVST	Q91bb5	spodoptera
45	47	43.5	922	2	Q5RKY4_MOUSE	Q5rky4	mus musculu

# ALIGNMENTS

## RESULT 1

Q8ISB3\_9NEOP

ID Q8ISB3\_9NEOP PRELIMINARY; PRT; 507 AA.

AC Q8ISB3;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Fibroin (Fragment).  
 OS Antheraea mylitta.  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;  
 OC Saturniidae; Saturniinae; Saturniini; Antheraea.  
 OX NCBI\_TaxID=34739;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Datta A., Ghosh A.K., Kundu S.C.;  
 RT "Differential expression of the fibroin gene in developmental stages  
 RT of silkworm, Antheraea mylitta (Saturniidae).";  
 RL Comp. Biochem. Physiol. B, Comp. Biochem. 129:197-204 (2001).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Datta A., Ghosh A.K., Kundu S.C.;  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AY136274; AAN28165.1; -; Genomic\_DNA.  
 FT NON\_TER 507 507  
 SQ SEQUENCE 507 AA; 45401 MW; C4FF5F3937268386 CRC64;

Query Match 100.0%; Score 108; DB 2; Length 507;  
 Best Local Similarity 100.0%; Pred. No. 4.2e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VETIVLEEDPYGHEDIYEED 20  
 |||||  
 Db 60 VETIVLEEDPYGHEDIYEED 79

Search completed: December 16, 2005, 23:52:14  
 Job time : 129.857 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2005, 23:08:17 ; Search time 81.25 Seconds  
(without alignments)  
70.301 Million cell updates/sec

Title: US-10-789-494B-8  
Perfect score: 72  
Sequence: 1 DDGFVLDGGYDSE 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_21:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*  
9: geneseqp2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	72	100.0	13	8	ADU51212	Adu51212 Silkworm
2	72	100.0	13	8	ADU51238	Adu51238 Gut silkw
3	72	100.0	28	8	ADU51203	Adu51203 Gut silkw
4	72	100.0	2655	7	ADO59401	Ado59401 Antheraea
5	45	62.5	138	3	AAG19475	Aag19475 Arabidops
6	45	62.5	138	3	AAG42768	Aag42768 Arabidops
7	45	62.5	157	3	AAG19474	Aag19474 Arabidops
8	45	62.5	157	3	AAG42767	Aag42767 Arabidops

9	45	62.5	173	3	AAG42766	Aag42766	Arabidops
10	45	62.5	173	3	AAG19473	Aag19473	Arabidops
11	42	58.3	194	6	ABP75555	Abp75555	Human sec
12	42	58.3	309	8	ADS16685	Ads16685	Bartonell
13	42	58.3	344	8	ADT59554	Adt59554	Plant pol
14	42	58.3	350	3	AAG27315	Aag27315	Arabidops
15	42	58.3	358	3	AAG27509	Aag27509	Arabidops
16	42	58.3	647	3	AAG37772	Aag37772	Arabidops
17	42	58.3	647	3	AAG20910	Aag20910	Arabidops
18	42	58.3	853	8	ADS24819	Ads24819	Bacterial
19	41	56.9	252	8	ADS42066	Ads42066	Bacterial
20	41	56.9	360	8	ADI39309	Adi39309	S. hygros
21	41	56.9	499	9	ADW17813	Adw17813	Pinus rad
22	41	56.9	499	9	ADW18484	Adw18484	Pinus rad
23	41	56.9	508	9	ADW17814	Adw17814	Pinus rad
24	41	56.9	632	2	AAV36992	Aay36992	Protein i
25	41	56.9	857	6	ABU40019	Abu40019	Protein e
26	40.5	56.2	929	8	ADH74699	Adh74699	Rice blas
27	40	55.6	53	4	ABG24441	Abg24441	Novel hum
28	40	55.6	72	5	ABP06134	Abp06134	Human ORF
29	40	55.6	115	3	AAV66693	Aay66693	Membrane-
30	40	55.6	115	4	AAU29107	Aau29107	Human PRO
31	40	55.6	115	4	AAB65216	Aab65216	Human PRO
32	40	55.6	115	6	ABU58483	Abu58483	Human PRO
33	40	55.6	115	6	ABU88031	Abu88031	Novel hum
34	40	55.6	115	6	ABU84346	Abu84346	Human sec
35	40	55.6	115	6	ABR66220	Abr66220	Human sec
36	40	55.6	115	6	ABR65610	Abr65610	Human sec
37	40	55.6	115	6	ABU99550	Abu99550	Human sec
38	40	55.6	115	6	ABU58031	Abu58031	Human PRO
39	40	55.6	115	6	ABU59109	Abu59109	Novel hum
40	40	55.6	115	6	ABU82621	Abu82621	Human sec
41	40	55.6	115	6	ABU82789	Abu82789	Human PRO
42	40	55.6	115	6	ABU89910	Abu89910	Novel hum
43	40	55.6	115	6	ABR68159	Abr68159	Human sec
44	40	55.6	115	6	ABU60540	Abu60540	Human sec
45	40	55.6	115	6	ABU96212	Abu96212	Novel hum

# ALIGNMENTS

## RESULT 1

ADU51212

ID ADU51212 standard; peptide; 13 AA.

XX

AC ADU51212;

XX

DT 24-FEB-2005 (first entry)

XX

DE Silkworm fibroin-derived fibroblast proliferation peptide 9.

XX

KW vulnerary; cell proliferation; wound healing; cell adhesion; cosmetics;

KW cell culture; fibroin.

XX

OS Bombycoidea.

OS Synthetic.

XX  
 PN JP2004339189-A.  
 XX  
 PD 02-DEC-2004.  
 XX  
 PF 04-DEC-2003; 2003JP-00406608.  
 XX  
 PR 28-FEB-2003; 2003JP-00055048.  
 XX  
 PA (DOKU-) DOKURITSU GYOSEI HOJIN NOGYO SEIBUTSU SH.  
 PA (TSUB/) TSUBOUCHI K.  
 XX  
 DR WPI; 2004-827614/82.  
 XX  
 PT New peptide having excellent cell growth promoting activity, for use as a  
 PT cell growth promoter, cell adhesion agent, wound healing-promoting agent,  
 PT cosmetic and cell culture base material.  
 XX  
 PS Claim 2; Page; 27pp; Japanese.  
 XX  
 CC The invention relates to a novel peptide having excellent cell growth  
 CC promoting activity. The peptide of the invention demonstrates vulnerary  
 CC activity and may be utilised as a cell growth promoter, cell adhesion  
 CC agent, wound healing-promoting agent or cosmetic and cell culture base  
 CC material. The current sequence is that of a silkworm fibroin-derived  
 CC fibroblast proliferation peptide of the invention.  
 XX  
 SQ Sequence 13 AA;

Query Match 100.0%; Score 72; DB 8; Length 13;  
 Best Local Similarity 100.0%; Pred. No. 4.8e-05;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDGFVLDGGYDSE 13  
 |||||  
 Db 1 DDGFVLDGGYDSE 13

# RESULT 5

AAG19475

ID AAG19475 standard; protein; 138 AA.

XX

AC AAG19475;

XX

DT 17-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 21292.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-00301439.  
XX  
PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.

PR	18-JUN-1999;	99US-0139461P.
PR	18-JUN-1999;	99US-0139462P.
PR	18-JUN-1999;	99US-0139463P.
PR	18-JUN-1999;	99US-0139750P.
PR	18-JUN-1999;	99US-0139763P.
PR	21-JUN-1999;	99US-0139817P.
PR	22-JUN-1999;	99US-0139899P.
PR	23-JUN-1999;	99US-0140353P.
PR	23-JUN-1999;	99US-0140354P.
PR	24-JUN-1999;	99US-0140695P.
PR	28-JUN-1999;	99US-0140823P.
PR	29-JUN-1999;	99US-0140991P.
PR	30-JUN-1999;	99US-0141287P.
PR	01-JUL-1999;	99US-0141842P.
PR	01-JUL-1999;	99US-0142154P.
PR	02-JUL-1999;	99US-0142055P.
PR	06-JUL-1999;	99US-0142390P.
PR	08-JUL-1999;	99US-0142803P.
PR	09-JUL-1999;	99US-0142920P.
PR	12-JUL-1999;	99US-0142977P.
PR	13-JUL-1999;	99US-0143542P.
PR	14-JUL-1999;	99US-0143624P.
PR	15-JUL-1999;	99US-0144005P.
PR	16-JUL-1999;	99US-0144085P.
PR	16-JUL-1999;	99US-0144086P.
PR	19-JUL-1999;	99US-0144325P.
PR	19-JUL-1999;	99US-0144331P.
PR	19-JUL-1999;	99US-0144332P.
PR	19-JUL-1999;	99US-0144333P.
PR	19-JUL-1999;	99US-0144334P.
PR	19-JUL-1999;	99US-0144335P.
PR	20-JUL-1999;	99US-0144352P.
PR	20-JUL-1999;	99US-0144632P.
PR	20-JUL-1999;	99US-0144884P.
PR	21-JUL-1999;	99US-0144814P.
PR	21-JUL-1999;	99US-0145086P.
PR	21-JUL-1999;	99US-0145088P.
PR	22-JUL-1999;	99US-0145085P.
PR	22-JUL-1999;	99US-0145087P.
PR	22-JUL-1999;	99US-0145089P.
PR	22-JUL-1999;	99US-0145192P.
PR	23-JUL-1999;	99US-0145145P.
PR	23-JUL-1999;	99US-0145218P.
PR	23-JUL-1999;	99US-0145224P.
PR	26-JUL-1999;	99US-0145276P.
PR	27-JUL-1999;	99US-0145913P.
PR	27-JUL-1999;	99US-0145918P.
PR	27-JUL-1999;	99US-0145919P.
PR	28-JUL-1999;	99US-0145951P.
PR	02-AUG-1999;	99US-0146386P.
PR	02-AUG-1999;	99US-0146388P.
PR	02-AUG-1999;	99US-0146389P.
PR	03-AUG-1999;	99US-0147038P.
PR	04-AUG-1999;	99US-0147204P.
PR	04-AUG-1999;	99US-0147302P.
PR	05-AUG-1999;	99US-0147192P.
PR	05-AUG-1999;	99US-0147260P.

PR	06-AUG-1999;	99US-0147303P.
PR	06-AUG-1999;	99US-0147416P.
PR	09-AUG-1999;	99US-0147493P.
PR	09-AUG-1999;	99US-0147935P.
PR	10-AUG-1999;	99US-0148171P.
PR	11-AUG-1999;	99US-0148319P.
PR	12-AUG-1999;	99US-0148341P.
PR	13-AUG-1999;	99US-0148565P.
PR	13-AUG-1999;	99US-0148684P.
PR	16-AUG-1999;	99US-0149368P.
PR	17-AUG-1999;	99US-0149175P.
PR	18-AUG-1999;	99US-0149426P.
PR	20-AUG-1999;	99US-0149722P.
PR	20-AUG-1999;	99US-0149723P.
PR	20-AUG-1999;	99US-0149929P.
PR	23-AUG-1999;	99US-0149902P.
PR	23-AUG-1999;	99US-0149930P.
PR	25-AUG-1999;	99US-0150566P.
PR	26-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151066P.
PR	27-AUG-1999;	99US-0151080P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156458P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.

PR 22-OCT-1999; 99US-0160980P.  
 PR 22-OCT-1999; 99US-0160981P.  
 PR 22-OCT-1999; 99US-0160989P.  
 PR 25-OCT-1999; 99US-0161404P.  
 PR 25-OCT-1999; 99US-0161405P.  
 PR 25-OCT-1999; 99US-0161406P.  
 PR 26-OCT-1999; 99US-0161359P.  
 PR 26-OCT-1999; 99US-0161360P.  
 PR 26-OCT-1999; 99US-0161361P.  
 PR 28-OCT-1999; 99US-0161920P.  
 PR 28-OCT-1999; 99US-0161992P.  
 PR 28-OCT-1999; 99US-0161993P.  
 PR 29-OCT-1999; 99US-0162142P.

Query Match 62.5%; Score 45; DB 3; Length 138;  
 Best Local Similarity 63.6%; Pred. No. 18;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DDGFVLDGGYD 11  
 ||| ::|||:  
 Db 40 DDGELIDGGFD 50

Search completed: December 16, 2005, 23:46:13  
 Job time : 84.25 secs

GenCore version 5.1.6  
 Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2005, 23:40:43 ; Search time 19.7321 Seconds  
 (without alignments)  
 54.469 Million cell updates/sec

Title: US-10-789-494B-8  
 Perfect score: 72  
 Sequence: 1 DDGFVLDGGYDSE 13

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
 1: /cgn2\_6/ptodata/1/iaa/5\_COMB.pep:\*

2: /cgn2\_6/ptodata/1/iaa/6\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/H\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/RE\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query		DB	ID	Description
		% Match	Length			
1	40	55.6	115	2	US-09-991-181-227	Sequence 227, App
2	40	55.6	115	2	US-09-990-444-227	Sequence 227, App
3	40	55.6	115	2	US-09-997-333-227	Sequence 227, App
4	40	55.6	115	2	US-09-992-598-227	Sequence 227, App
5	40	55.6	243	2	US-09-134-000C-4106	Sequence 4106, Ap
6	40	55.6	459	2	US-09-303-518D-214	Sequence 214, App
7	40	55.6	620	2	US-09-248-796A-20646	Sequence 20646, A
8	39	54.2	100	1	US-08-464-517-11	Sequence 11, Appl
9	39	54.2	100	1	US-08-246-361A-11	Sequence 11, Appl
10	39	54.2	100	2	US-08-463-772-11	Sequence 11, Appl
11	39	54.2	100	4	PCT-US93-05000-11	Sequence 11, Appl
12	39	54.2	148	1	US-08-460-694-7	Sequence 7, Appli
13	39	54.2	148	2	US-08-460-744-7	Sequence 7, Appli
14	39	54.2	148	2	US-07-667-711B-7	Sequence 7, Appli
15	39	54.2	159	2	US-09-270-767-32900	Sequence 32900, A
16	39	54.2	159	2	US-09-270-767-48117	Sequence 48117, A
17	39	54.2	318	2	US-09-120-051D-42	Sequence 42, Appl
18	38	52.8	16	2	US-08-822-322-4	Sequence 4, Appli
19	38	52.8	16	2	US-09-466-109-4	Sequence 4, Appli
20	38	52.8	44	2	US-10-241-602B-11	Sequence 11, Appl
21	38	52.8	126	2	US-09-471-276-1485	Sequence 1485, Ap
22	38	52.8	173	2	US-09-311-021-132	Sequence 132, App
23	38	52.8	174	2	US-09-149-476-424	Sequence 424, App
24	38	52.8	192	2	US-09-248-796A-17957	Sequence 17957, A
25	38	52.8	251	2	US-08-822-322-9	Sequence 9, Appli
26	38	52.8	251	2	US-09-466-109-9	Sequence 9, Appli
27	38	52.8	252	2	US-08-822-322-8	Sequence 8, Appli
28	38	52.8	252	2	US-09-466-109-8	Sequence 8, Appli
29	38	52.8	290	2	US-09-248-796A-18749	Sequence 18749, A
30	38	52.8	507	2	US-09-198-452A-174	Sequence 174, App
31	38	52.8	536	2	US-09-252-991A-33121	Sequence 33121, A
32	38	52.8	714	2	US-09-248-796A-18753	Sequence 18753, A
33	38	52.8	816	1	US-08-820-170A-37	Sequence 37, Appl
34	38	52.8	816	2	US-09-055-699-37	Sequence 37, Appl
35	38	52.8	816	2	US-09-273-565-37	Sequence 37, Appl
36	38	52.8	816	2	US-09-565-538-37	Sequence 37, Appl
37	38	52.8	816	2	US-09-661-468-37	Sequence 37, Appl
38	38	52.8	816	2	US-09-976-165-37	Sequence 37, Appl
39	38	52.8	821	2	US-09-438-185A-155	Sequence 155, App
40	38	52.8	1098	2	US-09-602-787A-48	Sequence 48, Appl
41	38	52.8	1874	2	US-09-602-787A-46	Sequence 46, Appl
42	38	52.8	5069	2	US-10-042-665A-5	Sequence 5, Appli



43	37	51.4	51	1	US-08-870-518-24
44	37	51.4	115	1	US-08-542-363-39
45	37	51.4	115	2	US-09-100-089-39

Sequence 24, Appl  
Sequence 39, Appl  
Sequence 39, Appl

# ALIGNMENTS

## RESULT 1

US-09-991-181-227

; Sequence 227, Application US/09991181

; Patent No. 6913919

; GENERAL INFORMATION:

; APPLICANT: Ashkenazi, Avi J.

; APPLICANT: Baker, Kevin P.

; APPLICANT: Botstein, David

; APPLICANT: Desnoyers, Luc

; APPLICANT: Eaton, Dan L.

; APPLICANT: Ferrara, Napoleone

; APPLICANT: Fong, Sherman

; APPLICANT: Gerber, Hanspeter

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Grimaldi, J. Christopher

; APPLICANT: Gurney, Austin L.

; APPLICANT: Kljavin, Ivar J.

; APPLICANT: Napier, Mary A.

; APPLICANT: Pan, James

; APPLICANT: Paoni, Nicholas F.

; APPLICANT: Roy, Margaret Ann

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K.

; APPLICANT: Williams, P. Mickey

; APPLICANT: Wood, William I.

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic

; TITLE OF INVENTION: Acids Encoding the Same

; FILE REFERENCE: P2730P1C53

; CURRENT APPLICATION NUMBER: US/09/991,181

; CURRENT FILING DATE: 2001-11-16

; PRIOR APPLICATION NUMBER: 60/049787

; PRIOR FILING DATE: 1997-06-16

; PRIOR APPLICATION NUMBER: 60/062250

; PRIOR FILING DATE: 1997-10-17

; PRIOR APPLICATION NUMBER: 60/065186

; PRIOR FILING DATE: 1997-11-12

; PRIOR APPLICATION NUMBER: 60/065311

; PRIOR FILING DATE: 1997-11-13

; PRIOR APPLICATION NUMBER: 60/066770

; PRIOR FILING DATE: 1997-11-24

; PRIOR APPLICATION NUMBER: 60/075945

; PRIOR FILING DATE: 1998-02-25

; PRIOR APPLICATION NUMBER: 60/078910

; PRIOR FILING DATE: 1998-03-20

; PRIOR APPLICATION NUMBER: 60/083322

; PRIOR FILING DATE: 1998-04-28  
; PRIOR APPLICATION NUMBER: 60/084600  
; PRIOR FILING DATE: 1998-05-07  
; PRIOR APPLICATION NUMBER: 60/087106  
; PRIOR FILING DATE: 1998-05-28  
; PRIOR APPLICATION NUMBER: 60/087607  
; PRIOR FILING DATE: 1998-06-02  
; PRIOR APPLICATION NUMBER: 60/087609  
; PRIOR FILING DATE: 1998-06-02  
; PRIOR APPLICATION NUMBER: 60/087759  
; PRIOR FILING DATE: 1998-06-02  
; PRIOR APPLICATION NUMBER: 60/087827  
; PRIOR FILING DATE: 1998-06-03  
; PRIOR APPLICATION NUMBER: 60/088021  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088025  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088026  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088028  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088029  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088030  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088033  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088326  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088167  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088202  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088212  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088217  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088655  
; PRIOR FILING DATE: 1998-06-09  
; PRIOR APPLICATION NUMBER: 60/088734  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088738  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088742  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088810  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088824  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088826  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088858  
; PRIOR FILING DATE: 1998-06-11  
; PRIOR APPLICATION NUMBER: 60/088861  
; PRIOR FILING DATE: 1998-06-11  
; PRIOR APPLICATION NUMBER: 60/088876  
; PRIOR FILING DATE: 1998-06-11

; PRIOR APPLICATION NUMBER: 60/089105  
; PRIOR FILING DATE: 1998-06-12  
; PRIOR APPLICATION NUMBER: 60/089440  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/089512  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/089514  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/089532  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089538  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089598  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089599  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089600  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089653  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089801  
; PRIOR FILING DATE: 1998-06-18  
; PRIOR APPLICATION NUMBER: 60/089907  
; PRIOR FILING DATE: 1998-06-18  
; PRIOR APPLICATION NUMBER: 60/089908  
; PRIOR FILING DATE: 1998-06-18  
; PRIOR APPLICATION NUMBER: 60/089947  
; PRIOR FILING DATE: 1998-06-19  
; PRIOR APPLICATION NUMBER: 60/089948  
; PRIOR FILING DATE: 1998-06-19  
; PRIOR APPLICATION NUMBER: 60/089952  
; PRIOR FILING DATE: 1998-06-19  
; PRIOR APPLICATION NUMBER: 60/090246  
; PRIOR FILING DATE: 1998-06-22  
; PRIOR APPLICATION NUMBER: 60/090252  
; PRIOR FILING DATE: 1998-06-22  
; PRIOR APPLICATION NUMBER: 60/090254  
; PRIOR FILING DATE: 1998-06-22  
; PRIOR APPLICATION NUMBER: 60/090349  
; PRIOR FILING DATE: 1998-06-23  
; PRIOR APPLICATION NUMBER: 60/090355  
; PRIOR FILING DATE: 1998-06-23  
; PRIOR APPLICATION NUMBER: 60/090429  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090431  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090435  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090444  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090445  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090472  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090535  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090540

```

; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

```

```

Query Match          55.6%; Score 40; DB 2; Length 115;
Best Local Similarity 54.5%; Pred. No. 25;
Matches      6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      2 DGFVLDGGYDS 12
      :|| ::|||:|
Db     81 NGFYINGGYES 91

```

Search completed: December 17, 2005, 00:12:08  
Job time : 20.7321 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2005, 23:52:28 ; Search time 68.7143 Seconds  
(without alignments)  
79.049 Million cell updates/sec

Title: US-10-789-494B-8  
Perfect score: 72  
Sequence: 1 DDGFVLDGGYDSE 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA\_Main:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	72	100.0	13	5	US-10-789-494B-8	Sequence 8, Appli
2	72	100.0	13	5	US-10-789-494B-68	Sequence 68, Appl
3	72	100.0	28	5	US-10-789-494B-55	Sequence 55, Appl
4	46	63.9	166	4	US-10-424-599-206658	Sequence 206658,
5	43	59.7	173	4	US-10-424-599-265404	Sequence 265404,
6	43	59.7	478	4	US-10-156-761-10964	Sequence 10964, A
7	42	58.3	174	4	US-10-767-701-34455	Sequence 34455, A
8	42	58.3	174	4	US-10-425-115-357755	Sequence 357755,
9	42	58.3	174	4	US-10-425-115-357760	Sequence 357760,
10	42	58.3	296	4	US-10-767-701-38933	Sequence 38933, A
11	42	58.3	344	5	US-10-739-930-9631	Sequence 9631, Ap
12	42	58.3	853	4	US-10-369-493-13852	Sequence 13852, A
13	42	58.3	3649	5	US-10-732-923-20588	Sequence 20588, A
14	41	56.9	86	4	US-10-156-761-9910	Sequence 9910, Ap
15	41	56.9	91	4	US-10-424-599-188864	Sequence 188864,
16	41	56.9	252	4	US-10-369-493-20496	Sequence 20496, A

17	41	56.9	342	4	US-10-437-963-117472	Sequence 117472,
18	41	56.9	360	4	US-10-461-194-151	Sequence 151, App
19	41	56.9	819	5	US-10-503-135-116	Sequence 116, App
20	41	56.9	857	4	US-10-282-122A-67943	Sequence 67943, A
21	40.5	56.2	236	4	US-10-425-115-278601	Sequence 278601,
22	40.5	56.2	929	4	US-10-436-323-3	Sequence 3, Appli
23	40	55.6	53	5	US-10-450-763-54800	Sequence 54800, A
24	40	55.6	109	4	US-10-424-599-242651	Sequence 242651,
25	40	55.6	111	4	US-10-767-701-38775	Sequence 38775, A
26	40	55.6	115	3	US-09-989-722-227	Sequence 227, App
27	40	55.6	115	3	US-09-989-723-227	Sequence 227, App
28	40	55.6	115	3	US-09-989-279-227	Sequence 227, App
29	40	55.6	115	3	US-09-989-727-227	Sequence 227, App
30	40	55.6	115	3	US-09-989-731-227	Sequence 227, App
31	40	55.6	115	3	US-09-989-732-227	Sequence 227, App
32	40	55.6	115	3	US-09-991-073-227	Sequence 227, App
33	40	55.6	115	3	US-09-990-442-227	Sequence 227, App
34	40	55.6	115	3	US-09-991-163-227	Sequence 227, App
35	40	55.6	115	3	US-09-993-604-227	Sequence 227, App
36	40	55.6	115	3	US-09-990-456-227	Sequence 227, App
37	40	55.6	115	3	US-09-989-721-227	Sequence 227, App
38	40	55.6	115	3	US-09-992-598-227	Sequence 227, App
39	40	55.6	115	3	US-09-989-293A-227	Sequence 227, App
40	40	55.6	115	3	US-09-989-735-227	Sequence 227, App
41	40	55.6	115	3	US-09-990-444-227	Sequence 227, App
42	40	55.6	115	3	US-09-991-181-227	Sequence 227, App
43	40	55.6	115	3	US-09-989-730-227	Sequence 227, App
44	40	55.6	115	3	US-09-990-436-227	Sequence 227, App
45	40	55.6	115	3	US-09-993-687-227	Sequence 227, App

#### ALIGNMENTS

##### RESULT 1

US-10-789-494B-8

; Sequence 8, Application US/10789494B

; Publication No. US20050143296A1

; GENERAL INFORMATION:

; APPLICANT: TSUBOUCHI, Kozo

; APPLICANT: YAMADA, Hiromi

; TITLE OF INVENTION: EXTRACTION AND UTILIZATION OF CELL

; TITLE OF INVENTION: GROWTH-PROMOTING PEPTIDES FROM SILK PROTEIN

; FILE REFERENCE: OPS 635

; CURRENT APPLICATION NUMBER: US/10/789,494B

; CURRENT FILING DATE: 2004-02-27

; PRIOR APPLICATION NUMBER: JP 2003-55048

; PRIOR FILING DATE: 2003-02-28

; NUMBER OF SEQ ID NOS: 85

; SEQ ID NO 8

; LENGTH: 13

; TYPE: PRT

; ORGANISM: Antheraea yamamai

US-10-789-494B-8

Query Match 100.0%; Score 72; DB 5; Length 13;

Best Local Similarity 100.0%; Pred. No. 4.4e-05;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDGFVLDGGYDSE 13  
|||||||  
Db 1 DDGFVLDGGYDSE 13

RESULT 4

US-10-424-599-206658

; Sequence 206658, Application US/10424599

; Publication No. US20040031072A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 206658

; LENGTH: 166

; TYPE: PRT

; ORGANISM: Glycine max

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_2863C.1.pep

US-10-424-599-206658

Query Match 63.9%; Score 46; DB 4; Length 166;

Best Local Similarity 61.5%; Pred. No. 13;

Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DDGFVLDGGYDSE 13  
||| | :||::||  
Db 125 DDGVVEEGGFESE 137

Search completed: December 17, 2005, 00:17:34

Job time : 69.7143 secs

GenCore version 5.1.6

Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2005, 23:53:24 ; Search time 3.48214 Seconds  
(without alignments)  
25.151 Million cell updates/sec

Title: US-10-789-494B-8

Perfect score: 72

Sequence: 1 DDGFVLDGGYDSE 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 51470 seqs, 6736768 residues

Total number of hits satisfying chosen parameters: 51470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA\_New:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	% Query		Length	DB	ID	Description
	Score	Match				
1	40	55.6	473	6	US-10-467-657-486	Sequence 486, App
2	38	52.8	816	7	US-11-090-439-48	Sequence 48, Appl
3	37	51.4	95	6	US-10-467-657-8782	Sequence 8782, Ap
4	35	48.6	88	7	US-11-053-076-179	Sequence 179, App
5	35	48.6	94	7	US-11-053-076-184	Sequence 184, App
6	35	48.6	348	6	US-10-878-556A-111	Sequence 111, App
7	35	48.6	713	6	US-10-467-657-1012	Sequence 1012, Ap
8	35	48.6	713	7	US-11-190-799-2	Sequence 2, Appli
9	35	48.6	713	7	US-11-190-799-4	Sequence 4, Appli
10	35	48.6	1267	7	US-11-109-156-35	Sequence 35, Appl
11	34	47.2	192	6	US-10-467-657-6808	Sequence 6808, Ap
12	34	47.2	376	7	US-11-055-822-152	Sequence 152, App
13	34	47.2	507	6	US-10-467-657-1612	Sequence 1612, Ap
14	34	47.2	1044	7	US-11-091-668-2	Sequence 2, Appli
15	34	47.2	1313	7	US-11-091-668-4	Sequence 4, Appli
16	34	47.2	7102	7	US-11-143-980-48	Sequence 48, Appl
17	34	47.2	7968	7	US-11-143-980-49	Sequence 49, Appl
18	33	45.8	124	6	US-10-467-657-3838	Sequence 3838, Ap
19	33	45.8	133	6	US-10-467-657-38	Sequence 38, Appl
20	33	45.8	133	6	US-10-467-657-8452	Sequence 8452, Ap
21	33	45.8	257	6	US-10-467-657-7456	Sequence 7456, Ap
22	33	45.8	282	6	US-10-878-556A-125	Sequence 125, App



23	33	45.8	430	6	US-10-793-626-130	Sequence 130, App
24	33	45.8	436	7	US-11-116-939-9	Sequence 9, Appli
25	33	45.8	466	6	US-10-467-657-2360	Sequence 2360, Ap
26	33	45.8	471	6	US-10-995-561-901	Sequence 901, App
27	33	45.8	678	7	US-11-102-240-34	Sequence 34, Appl
28	33	45.8	782	6	US-10-821-234-1592	Sequence 1592, Ap
29	33	45.8	824	7	US-11-116-939-11	Sequence 11, Appl
30	32.5	45.1	423	6	US-10-467-962B-85	Sequence 85, Appl
31	32	44.4	35	6	US-10-467-657-7098	Sequence 7098, Ap
32	32	44.4	35	6	US-10-467-657-8030	Sequence 8030, Ap
33	32	44.4	78	7	US-11-000-463-469	Sequence 469, App
34	32	44.4	78	7	US-11-000-463-941	Sequence 941, App
35	32	44.4	125	6	US-10-467-657-5290	Sequence 5290, Ap
36	32	44.4	155	6	US-10-467-657-7322	Sequence 7322, Ap
37	32	44.4	158	6	US-10-467-657-7326	Sequence 7326, Ap
38	32	44.4	227	6	US-10-467-657-970	Sequence 970, App
39	32	44.4	263	6	US-10-467-657-284	Sequence 284, App
40	32	44.4	263	6	US-10-467-657-5298	Sequence 5298, Ap
41	32	44.4	322	6	US-10-793-626-1912	Sequence 1912, Ap
42	32	44.4	334	7	US-11-055-822-24	Sequence 24, Appl
43	32	44.4	348	7	US-11-102-497-3	Sequence 3, Appli
44	32	44.4	348	7	US-11-102-497-11	Sequence 11, Appl
45	32	44.4	360	6	US-10-467-657-5376	Sequence 5376, Ap

#### ALIGNMENTS

##### RESULT 1

US-10-467-657-486

```
; Sequence 486, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 486
; LENGTH: 473
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-486
```

```
Query Match          55.6%; Score 40; DB 6; Length 473;
Best Local Similarity 77.8%; Pred. No. 9.9;
Matches      7; Conservative      0; Mismatches      2; Indels      0; Gaps      0;
```

Qy 1 DDGFVLDGG 9

Db                    417 DDGFCADGG 425

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

Result No.	Score	Query			ID	Description
		Match	Length	DB		
1	72	100.0	2639	2	T31328	fibroin - Chinese
2	45	62.5	173	2	F84630	hypothetical prote
3	44	61.1	546	2	A69484	hypothetical prote
4	44	61.1	938	2	AF1772	internalin-like pr
5	43	59.7	657	2	B84791	hypothetical prote
6	42	58.3	223	2	D84233	hypothetical prote
7	42	58.3	249	2	G87105	probable amidotran

8	42	58.3	867	2	G69485	DNA-directed RNA p
9	41	56.9	819	2	C71544	leucine-tRNA ligas
10	41	56.9	1196	2	A29130	beta-amylase (EC 3
11	40	55.6	42	2	F81852	hypothetical prote
12	40	55.6	520	2	I51556	recombination acti
13	40	55.6	706	2	S53035	probable lysosphosp
14	39	54.2	217	2	G85356	hypothetical prote
15	39	54.2	329	1	G69210	conserved hypothet
16	39	54.2	331	1	C69026	acetylpolyamine am
17	39	54.2	381	2	AG0110	probable exported
18	39	54.2	482	2	A34948	cyclin-related cel
19	39	54.2	482	2	S01153	cell division cont
20	39	54.2	499	2	AI1107	internalin E [impo
21	39	54.2	528	2	S42510	Rag-2 protein - ch
22	39	54.2	686	2	S43562	K08E5.3 protein -
23	39	54.2	813	2	G96494	protein F7F22.4 [i
24	39	54.2	855	2	A53296	DNA mismatch repai
25	39	54.2	889	1	S47162	DNA-directed RNA p
26	39	54.2	938	2	A56731	chromatin assembly
27	38	52.8	99	2	T09998	12-oxophytodienoat
28	38	52.8	211	2	E87632	conserved hypothet
29	38	52.8	229	2	AC3419	transcription regu
30	38	52.8	261	2	G87535	regulatory protein
31	38	52.8	284	2	AE3368	probable thiosulfa
32	38	52.8	292	2	E90180	conserved hypothet
33	38	52.8	337	2	AB3440	acetylspemidine d
34	38	52.8	360	1	S59311	alcohol dehydrogen
35	38	52.8	407	2	T40582	hypothetical prote
36	38	52.8	428	2	G81362	3-phosphoshikimate
37	38	52.8	428	2	JC5338	3-phosphoshikimate
38	38	52.8	551	2	C89134	protein F25G6.6 [i
39	38	52.8	694	2	S10632	replication initia
40	38	52.8	820	2	A86510	leucyl tRNA synthe
41	38	52.8	820	2	C72113	leucine-tRNA ligas
42	38	52.8	821	2	AB1126	internalin, peptid
43	38	52.8	970	2	E70533	probable sulfatase
44	38	52.8	1070	2	AI0484	probable autotrans
45	38	52.8	5069	2	T17464	rifamycin polyketi

#### ALIGNMENTS

##### RESULT 1

T31328

fibroin - Chinese oak silkmoth

C;Species: *Antheraea pernyi* (Chinese oak silkmoth)

C;Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004

C;Accession: T31328

R;Sezutsu, H.; Tamura, T.; Yukuhiro, K.

submitted to the EMBL Data Library, August 1998

A;Description: Characterization of the full length fibroin gene of a wild silkworm, *Antheraea pernyi*.

A;Reference number: Z20995

A;Accession: T31328

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-2639 <SEZ>  
A;Cross-references: UNIPROT:O76786; UNIPARC:UPI0000078D8E; EMBL:AF083334;  
NID:g3450882; PID:g3450883; PIDN:AAC32606.1  
C;Genetics:  
A;Introns: 14/3

Query Match 100.0%; Score 72; DB 2; Length 2639;  
Best Local Similarity 100.0%; Pred. No. 0.0019;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDGFVLDGGYDSE 13  
|||  
Db 2579 DDGFVLDGGYDSE 2591

RESULT 2

F84630

hypothetical protein At2g23940 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004

C;Accession: F84630

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.;  
Fujii, C.Y.; Mason, T.M.; Bowman, C.L.; Barnstead, M.E.; Feldblyum, T.V.; Buell,  
C.R.; Ketchum, K.A.; Lee, J.J.; Ronning, C.M.; Koo, H.; Moffat, K.S.; Cronin,  
L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.J.; Gill, J.E.; Adams,  
M.D.; Carrera, A.J.; Creasy, T.H.; Goodman, H.M.; Somerville, C.R.; Copenhaver,  
G.P.; Preuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser,  
C.M.; Venter, J.C.

Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis  
thaliana.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: F84630

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-173 <STO>

A;Cross-references: UNIPROT:O82222; UNIPARC:UPI00000A5B51; GB:AE002093;

NID:g3738323; PIDN:AAC63664.1; GSPDB:GN00139

C;Genetics:

A;Gene: At2g23940

A;Map position: 2

Query Match 62.5%; Score 45; DB 2; Length 173;  
Best Local Similarity 63.6%; Pred. No. 3;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DDGFVLDGGYD 11  
||| :|||  
Db 75 DDGELIDGGFD 85

Search completed: December 16, 2005, 23:53:17  
Job time : 14.4643 secs

OM protein - protein search, using sw model

Run on: December 16, 2005, 23:33:02 ; Search time 83.1071 Seconds  
(without alignments)  
110.362 Million cell updates/sec

Title: US-10-789-494B-8  
Perfect score: 72  
Sequence: 1 DDGFVLDGGYDSE 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_05.80:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query	Match Length			
1	72	100.0	151	2	Q95VQ0_ANTYA	Q95vq0 antheraea y
2	72	100.0	436	2	Q967T8_ANTPE	Q967t8 antheraea p
3	72	100.0	2639	2	O76786_ANTPE	O76786 antheraea p
4	72	100.0	2655	2	Q964F4_ANTYA	Q964f4 antheraea y
5	71	98.6	421	2	Q93119_ANTPE	Q93119 antheraea p
6	49	68.1	685	2	Q7VDT2_PROMA	Q7vdt2 prochloroco
7	47	65.3	425	2	Q6D6K1_ERWCT	Q6d6k1 erwinia car
8	47	65.3	428	2	Q6D355_ERWCT	Q6d355 erwinia car
9	46	63.9	910	2	Q6B390_9PROT	Q6b390 uncultured
10	45	62.5	173	2	Q8GYK4_ARATH	Q8gyk4 arabidopsis
11	45	62.5	173	2	Q8LEX3_ARATH	Q8lex3 arabidopsis
12	45	62.5	173	2	O82222_ARATH	O82222 arabidopsis
13	44	61.1	392	2	Q4LYV7_9BURK	Q4lyv7 burkholderi
14	44	61.1	458	2	Q5WHT0_BACSK	Q5wht0 bacillus cl
15	44	61.1	546	2	O28405_ARCFU	O28405 archaeoglob
16	44	61.1	938	2	Q927R4_LISIN	Q927r4 listeria in
17	43	59.7	56	2	Q5VQN1_ORYSA	Q5vqn1 oryza sativ
18	43	59.7	238	2	Q5TPW5_ANOGA	Q5tpw5 anopheles g
19	43	59.7	267	2	Q7QLB3_ANOGA	Q7qlb3 anopheles g

20	43	59.7	390	2	Q6T1G6_NEOFS	Q6t1g6 neoceratodu
21	43	59.7	478	2	Q82HT1_STRAW	Q82ht1 streptomyce
22	43	59.7	490	2	Q90XH9_9SARC	Q90xh9 protopterus
23	43	59.7	601	2	Q7Q2X2_ANOGA	Q7q2x2 anopheles g
24	43	59.7	657	2	Q9ZUT5_ARATH	Q9zut5 arabidopsis
25	43	59.7	729	2	Q54RD7_DICDI	Q54rd7 dictyosteli
26	43	59.7	1228	2	Q4FYU1_LEIMA	Q4fyu1 leishmania
27	42	58.3	223	2	Q9HRC6_HALSA	Q9hrc6 halobacteri
28	42	58.3	249	2	Q9CBU9_MYCLE	Q9cbu9 mycobacteri
29	42	58.3	364	2	Q5SPG9_BRARE	Q5spg9 brachydanio
30	42	58.3	390	2	Q6T1G5_LEPPA	Q6t1g5 lepidosiren
31	42	58.3	421	2	Q90XI2_9CHON	Q90xi2 triakis sp.
32	42	58.3	459	2	Q5L207_GEOKA	Q5l207 geobacillus
33	42	58.3	486	2	Q90XI3_CHIPU	Q90xi3 chiloscylli
34	42	58.3	510	2	Q5PEW7_SALPA	Q5pew7 salmonella
35	42	58.3	511	1	GUAA_MYCPE	Q8ews9 mycoplasma
36	42	58.3	520	2	Q89YV0_BACTN	Q89yv0 bacteroides
37	42	58.3	586	2	Q98J75_RHILO	Q98j75 rhizobium l
38	42	58.3	626	2	Q8I1Z9_PLAF7	Q8ilz9 plasmodium
39	42	58.3	647	2	Q8LDD0_ARATH	Q8ldd0 arabidopsis
40	42	58.3	647	2	Q9FM96_ARATH	Q9fm96 arabidopsis
41	42	58.3	823	2	Q723P8_LISMF	Q723p8 listeria mo
42	42	58.3	866	2	Q7S7S5_NEUCR	Q7s7s5 neurospora
43	42	58.3	867	2	O28391_ARCFU	O28391 archaeoglob
44	42	58.3	881	2	Q9C261_NEUCR	Q9c261 neurospora
45	42	58.3	1529	2	Q4SHR3_TETNG	Q4shr3 tetraodon n

# ALIGNMENTS

## RESULT 1

Q95VQ0\_ANTYA

ID Q95VQ0\_ANTYA PRELIMINARY; PRT; 151 AA.

AC Q95VQ0;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Fibroin heavy chain (Fragment).

GN Name=Fib-H;

OS Antheraea yamamai (Japanese oak silkmoth).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;

OC Saturniidae; Saturniinae; Saturniini; Antheraea.

OX NCBI\_TaxID=7121;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RA Zurovec M., Yang C., Sehna F.;

RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.

RN [2]

RP NUCLEOTIDE SEQUENCE.

RA Fedic R.;

RL Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.

DR EMBL; AF410906; AAL02118.1; -; mRNA.

FT NON\_TER 1 1

SQ SEQUENCE 151 AA; 13629 MW; 596775A00040475A CRC64;

Query Match 100.0%; Score 72; DB 2; Length 151;  
Best Local Similarity 100.0%; Pred. No. 0.00052;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DDGFVLDGGYDSE 13  
|||  
Db 91 DDGFVLDGGYDSE 103

Search completed: December 16, 2005, 23:52:17  
Job time : 86.1071 secs